

GenCore version 4.5  
 Copyright (c) 1993 - 2000 Compugen Ltd.  
 OM protein - protein search, using sw model  
 Run on: July 24, 2002, 04:18:22 ; Search time 67.11 Seconds  
 (without alignments)  
 4542.051 Million cell updates/sec  
 Title: US-10-044-807-2  
 Perfect score: 9588  
 Sequence: 1 MECCRRATPGTLLFLAFL.....LKQLCSQFKSRCCGTGKA 1762

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5  
 Searched: 562222 seqs, 172994929 residues  
 Total number of hits satisfying chosen parameters: 562222  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries  
 Database : SPTREMBL19:  
 1: sp\_archaea:  
 2: sp\_bacteria:  
 3: sp\_fungi:  
 4: sp\_human:  
 5: sp\_invertebrate:  
 6: sp\_mammal:  
 7: sp\_mhc:  
 8: sp\_organelle:  
 9: sp\_phage:  
 10: sp\_plant:  
 11: sp\_rodent:  
 12: sp\_virus:  
 13: sp\_vertebrate:  
 14: sp\_unclassified:  
 15: sp\_rvirus:  
 16: sp\_bacteriaph:  
 17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	2907	30.3	525	4 Q96RW4	Q96RW4 homo sapien
2	2625.5	27.4	766	4 P82987	P82987 homo sapien
3	2283.5	23.8	417	4 Q9BXY3	Q9BXY3 homo sapien
4	1629	17.0	1023	4 Q9ULI7	Q9ULI7 homo sapien
5	1279.5	13.3	1014	5 Q95R33	Q95R33 drosophila
6	1145.5	11.9	2165	5 Q19791	Q19791 caenorhabdi
7	1088.5	11.4	1059	5 P90884	P90884 caenorhabdi
8	1019.5	10.6	1280	11 Q9EPX2	Q9EPX2 mus musculu
9	994	10.4	192	11 Q9CX59	Q9CX59 mus musculu
10	978	10.2	1235	4 Q95428	Q95428 homo sapien
11	887	9.3	2167	5 Q76840	Q76840 caenorhabdi
12	882	9.2	2174	5 Q9G8R0	Q9G8R0 drosophila
13	882	9.2	3060	5 Q9VAV4	Q9VAV4 drosophila
14	865	9.0	1572	5 Q49388	Q49388 haemochus
15	857.5	8.9	3198	5 Q9U8G8	Q9U8G8 manduca sex
16	848.5	8.8	761	6 Q951Q2	Q951Q2 macaca fasc

## SUMMARIES

## ALIGNMENTS

## RESULT 1

Q96RW4 ID Q96RW4 PRELIMINARY; PRT; 525 AA.  
 AC Q96RW4;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ADAM-TS RELATED PROTEIN 1.  
 GN ADAMTSR1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_taxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hirohata S., Apte S.S.;  
 RT "A novel member of ADAM-TS related gene, ADAM-TSRI (A Disintegrin-like  
 RT And Metalloproteinase domain with Thrombospondin type I modules  
 RT Related gene-1).";  
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF176313; AAK84170.1; -;  
 SQ SEQUENCE 525 AA; 58351 MW; 3F180C3CAAA7BA68 CRC64;

Query Match 30.3%; Score 2907; DB 4; Length 525;  
 Best Local Similarity 99.6%; Pred. No. 3.2e-204;  
 Matches 523; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MECCRRATPGTLLFLAFL.....LKQLCSQFKSRCCGTGKA 1762  
 DB 1 MECCRRATPGTLLFLAFL.....LKQLCSQFKSRCCGTGKA 1762  
 QY 61 LSSKSCGRNIRYTCNVDCPPAGDFRAQOC SAHNDVKHHGQFYEWLPVSNPDNPPCS 120  
 DB 61 LSSKSCGRNIRYTCNVDCPPAGDFRAQOC SAHNDVKHHGQFYEWLPVSNPDNPPCS 120  
 QY 121 LKQAKGTTTLVWELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180  
 DB 121 LKQAKGTTTLVWELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180

QY	181	DGSTRVLRGQYKSOISATKSDDTVAIPYCSRHRLVGLKGPDLHLYLETKTLQGTGKNS	240
Db	181	DGSTRVLRGQYKSOISATKSDDTVAIPYCSRHRLVGLKGPDLHLYLETKTLQGTGKNS	240
QY	241	LSSTGTFLVNSSVDFOKFPDKKEILRMAGPLTADFIKIRNSGSADSTVQFIYQPIIHR	300
Db	241	LSSTGTFLVNSSVDFOKFPDKKEILRMAGPLTADFIKIRNSGSADSTVQFIYQPIIHR	300
QY	301	WRETDFFPCSNCTGGGQYLSAECYDLRSNRVADQYCHYYPENIKPKLOECNLDPCP	360
Db	301	WRETDFFPCSNCTGGGQYLSAECYDLRSNRVADQYCHYYPENIKPKLOECNLDPCP	360
QY	361	ASDGYKQIMPYDLVHPLPRWEATPTWATSCSSCGGGIQSRVSCVEEDIQGHVTSVEEWC	420
Db	361	ASDGYKQIMPYDLVHPLPRWEATPTWATSCSSCGGGIQSRVSCVEEDIQGHVTSVEEWC	420
QY	421	MYTPKMPAQPCNTFDCPKWLAQNSPCTVTCGGGLRYRVVLCIDHRGMHTGGCSPKTKP	480
Db	421	MYTPKMPAQPCNTFDCPKWLAQNSPCTVTCGGGLRYRVVLCIDHRGMHTGGCSPKTKP	480
QY	481	HIKEECIVPTPCYKPKELPVEAKLPWFKAQLEEGAAVSEEPS	525
Db	481	HIKEECIVPTPCYKPKELPVEAKLPWFKAQLEEGAAVSEEPS	525
RESULT	2		
ID	P82987	PRELIMINARY;	PRT; 766 AA.
AC	P82987;		
DT	01-JUN-2001 (Tremblrel. 17, Created)		
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)		
DT	01-OCT-2001 (Tremblrel. 18, Last annotation update)		
DE	ADAM-TSL3 PRECURSOR (FRAGMENT).		
GN	ADAMTSL3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Hirohata S., Anand-Apte B., Seldin M., Apte S.;		
RT	"Punctin, a member of a new family with similarities to ADAM-TS		
RT	proteases, is a component of extracellular matrix of skeletal		
RT	muscle."		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.		
DR	EMBL; AF237652; AAK15041.1; --		
DR	InterPro; IPR000884; TSP1.		
DR	SMART; SM00209; TSP1; 6.		
DR	PROSITE; PS50092; TSP1; 1.		
KW	Repeat; Signal.		
FT	SIGNAL 1 26	POTENTIAL.	
FT	CHAIN 27 766	ADAM-TSL3.	
FT	DOMAIN 79 123	TSP TYPE-1 1.	
FT	DOMAIN 422 474	TSP TYPE-1 2.	
FT	DOMAIN 482 528	TSP TYPE-1 3.	
FT	DOMAIN 568 625	TSP TYPE-1 4.	
FT	DOMAIN 648 703	TSP TYPE-1 5.	
FT	DOMAIN 707 759	TSP TYPE-1 6.	
FT	NON_TER 766		
SQ	SEQUENCE 766 AA; 85987 MW; A42613E87AE91719 CRC64;		
Query Match	27.4%;	Score 2625.5; DB 4; Length 766;	
Best Local Similarity	62.5%;	Pred. No. 2.3e-183;	
Matches	448; Conservative 104; Mismatches 144; Indels 21; Gaps 4;		
QY	22	SSRTARSEEDRDLGWDAGWPWSECRTGGGASYSRLRCLSSKSCGRNIRYTCNSVDC	81
Db	64	TSRNTSRDEKDNWDAGWDWDCSRCTGGGASYSRLRCLTGRNCRGQNTIRYKTCNSHDC	123

Query Match 23.8%; Score 2283.5; DB 4; Length 417;  
Best Local Similarity 97.6%; Pred. No. 9.8e-159;  
Matches 407; Conservative 3; Mismatches 6; Indels 1; Gaps 1;



```
QY 1742 QLKLCQLSFKSRCCGTC 1759
      ||| :||| :|
Db 1003 HMLCSDLRYKQCCQSC 1020

RESULT 5
Q95R33 PRELIMINARY; PRT; 1014 AA.
AC Q95R33;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE GH19218P.
GN CG2131.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;
RU Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061825; AA127636.1;
SQ SEQUENCE 1014 AA; 112241 MW; 030A1645935D5360 CRC64;

Query Match 13.3%; Score 1279.5; DB 5; Length 1014;
Best Local Similarity 27.4%; Pred. No. 1.2e-84;
Matches 307; Conservative 139; Mismatches 343; Indels 333; Gaps 25;

QY 34 GLWDANGPWSCTRTGGGASYLRRLUSSKCEGRNTRYTCNVDCPPPPAGDFRAQOC 93
      ||| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
Db 76 GOWSSWDWSTCSTCDGIMHOMRRRCSPGSCRGESTRYRICNMQPC-PEQQDFRSQC 134

QY 94 SAHNDVHKGQFYELVPSNDPDCSLKQCAKGTTLV- 131
      ||| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
Db 135 SAYNDVPDGLYKWTTP-HYDYVEPCALTGRHPAHLVDIESRTGDNABEAHYDEQS 193

QY 132 --VELAPKVLDTGTCYTESLDMCISGLCQIVGCCDHLQSTYKEDNCGVCGDSTCRLLVR 189
      ||| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
Db 194 VIVQLSARVQDGTCTCRSGSLDMCIGKQCRVGCDDLKIGTKKIDGCGVCGDGNCS- 250

QY 190 GOYKSQLSATKSDDTVVAIPYGSRHRLVLKGPDPHLYLETKTLQGTGKENSLSSTGTLV 249

Db 251 ----- 250

QY 250 DNSSVDFQKFPDKETLRMAGPLTADFIYKIRNSGSDSTVQFIYQPIIHRWRETDFPC 309
      ||| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
Db 251 -----QPLFN-WEMAPMSQC 264

QY 310 SATCGGQQLTSAECYDLRSNRVADQYCHYYPENIKPKKLQECNLIDPCPASDGYKQIM 369
      ||| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
Db 265 SVTCGGYKMSRPICRNRLTNADVDTLCSV--INRPEASVEQCNTHSCP----- 312

QY 370 PYDLXHLPLRWEATPWTACSSCGGIGQSRVCSVEEDIQGHVTSVEWKCMYTPKMPITA 429
      ||| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
Db 313 -----PRWTDWSTCSRSLCCHGYRMYVVCAES-NGIKTRVADIMCR-TPKPTIQ 362

QY 430 QPCNIFDCPKLQAEWSPTCTVTCGGGLRVVLCIDHRGMHTGGCSPKTKPHKKECIVP 489
      ||| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
Db 363 EFCIIIECPHWEVDWTCGVSVCGGIQMRGVECKSTDGSLSAKCDPLTKPQSMQOCSTG 422

QY 490 TPC- 492
      |
Db 423 IHCGSLNKVGGTITVGSRSRLNRSERQDSSDAEDNEDENEDGDDVDDLESQDSTD 482
```

```
QY 493 -----YKPKELPVEA-----KLFWKQAELEBGAAYSEEPS 525
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 483 GEGLSYADQPLLYAHRTQSLNLEAPDEPRTHLMNGSNNNNNGEDSESGPSL--DPT 540

QY 526 FIPE-AWSACTVTCGVGTQVIRCOVLLSFSSQVADLPIDECBGPKPASQ-RACYAGPC 583
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 541 YIKDNEWSPCSVTCGEGIRRTYCKIFLEYSRTVATVYDLSLCEGKKPHDEVERCVDPC 600

QY 584 SGEIPEFPNDE---TDGLFGQDFDELIDWEYEGFTKCSBSCGGGVQAVVCSLUNKQTR 640
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 601 M-LPSHGFDQDFPRDSIKGVSEPGKTYVWREQYTCSCASCLGVEELIINCVRDNG 658

QY 641 EPAENLCVTSRRPPQLKSCNLDPCPARWEIGKWSKSLCGVGLQDFVCSHLLSRE 700
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 659 RVWSFLCSPETKPEARVTCNDRPCPRWYSDYTPCSCGIGIKTREVOCIEHVTG 718

QY 701 MNETVILADELCROPKSTVOACNRFNCPAPWPAQWPCSRTCGGGVQKREVLCQKMA 760
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 719 GDNVWVNSMCPQPPADRCQCNVLDGVRWEVGENSKCSHTCGVGFQDKRKECKQIMA 778

QY 761 DGSFLELPETFCASAKPACQACKKDDCPSEWLLSDWTECSTSCGEGTQTRSAICRMUK 820
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 779 QEHKTERPESMCPSAKPADKKPCNVKPCPE-----DPKPV----- 815

QY 821 TGLSTVWNSTLCPPLPFSSIPCMATCARPSTKHSPIHAAARKVYIOTRRQRKLH 880
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 816 ----QINNST-----HIOHDP-----AKKIT 833

QY 881 FVVGGFAYLLPKTAVVLCRRARRVRKPLITWEKDOHLLISHTVTVAPFGYKIHRLKPS 940
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 834 LKVGAGLVFFCTQYKICPKVRYNRTKWSKDKHKLQSRKRYKVKSGKALRILDIIFR 893

QY 941 DAGVYTCAGAPAREHFVTKLIGNKRLVARPLS-----PRSEEVLAGKGGPKALQT- 994
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 894 DAGVYSCHAGLSAELSI-----EVKAKPGQIELEQOETERLVRKESG-TEALTS 945

QY 995 -HKHONGIFSN--GSKAEKRGAAANPGSYDDLYSLLEQG 1032
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 946 DMKSADGTGTSTAVGNPNHVNVAQOHGRRRQOOSERLQNG 987

RESULT 6
Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F25H8.3 PROTEIN.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadaty S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DDAJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey I., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirov J., Kirov J., Kirov J., Kirov J.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
```





```

QY 1753 -----SRCCGTCCK 1761
Db 1886 WSGPGECKSKNCQ 1900

RESULT 7
P90884
ID P90884 PRELIMINARY; PRT; 1059 AA.
AC P90884;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F53B6.2 PROTEIN.
GN F53B6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA none;
RX MEDLINE=99069613; PubMed=9851916;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81086; CAB03121.1; .
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 8.
DR PROSITE; PS00092; TSP1; 2.
KW Immunoglobulin domain.
KW SEQUENCE 1059 AA; 117768 MW; 5B182A42E41C8597 CRC64;

Query Match 11.48; Score 1088.5; DB 5; Length 1059;
Best Local Similarity 19.38; Pred. No. 1.3e-70;
Matches 350; Conservative 188; Mismatches 446; Indels 827; Gaps 50;

QY 11 TLLFLAFLLSSRTARGEEDRGL-WDAWGPWSCSRTCCGGGASYSRLRCLSSKSCGR 69
Db 6 TWVFLFLVLLTASPHV-----DALSWAASFPWSCTKTCGGVSRQLRCLTSK-CSGE 58
QY 70 NIRYTCNVDCPPAGDFRAQCSAHNDVKHHGQFYEWLPVNSDPDNPCLSKCAQAGTT 129
Db 59 SVREKVCQAQKTESKSLARDTICGG-BEIVSRGQ-----CEVVCRSRLTG 103
QY 130 LVVELAPKVLDTGTCY--TESLDMCTISGLCOJVGCDHQLGSIWKEDNCGVNCGDGTCLRV 188
Db 104 --ANFLWRVDDGTQCAATSRAVCSKSGCQIVGCDGLSSSRFFDACCVCGGGQDTC--- 158
QY 189 RGQYKSQLSATKSDDTVAIPYGSRHRLVLPKDPDHLYLETKLQGTGKENSLSGTGFL 248
Db 159 ----- 158
QY 249 VDNSSVDFQKFPDKELLRMAGPLTADFIVKIRNSGSADSTVQFIYQPIIHRWRETDFFP 308
Db 159 -DNKG-----FIWKV-----SEETYA 173
QY 309 CSATC-----GGGYQLTSAB-----CYDLRSNRVADQYCHYYPENIKPKPLQECNL 356
Db 174 CASNCDLIVDSGAGRSASTSQPIVVCVNAITGRVPEKLC---ADKLRPKVEARPCPM 230
QY 357 DPCPADSGYKQIMPYDLYHPLRPWEATPWATCSSCGGGIQSRAYSCVE--EDIQGHVTS 414
Db 231 LICPS-----RWMAADWTECVPHCGEGTRKREYVCYQTAHNVTVH--- 270

```

```

QY 1466 EFSCLAQNEAGVLMOKASIVIQDYHWSVDRLATCSASCGNRGVQOPRLKCLLNSTEVNPA 1525
Db 808 -----DEKARQLRKY-----GNELV----- 823
QY 1526 HCAGKVRPAVQPIACNRDRCPKRWMTWSAC-TRSCG-GGVQTRRVTCOKLKASISTP 1583
Db 824 -----ARWLGHWSECRQKTYCHVAGYQARGISC-KVTFHGEIRN 861
QY 1584 VSNMCTQVAK-RPVDQACNOOLCVENAFSSWGQCNCP-CICPHLAVOHROVFCQTRDG 1641
Db 862 VDNSTICESLASVRPPEIRPCHREDPCPRWEASQWSECSQRCQVSSMLAQKRRNVTCRFNG 921
QY 1642 IITLSEQCSALPRPVSTONCRSEACSVHVRVSLWTLTLCATCGNYGFQSRRCVCHARNK 1701
Db 922 TSVLQHQCDITNRPATTMDCPNQNKAMRTSDMGSCSECTGGVQLRLLSVCWISSGR 981
QY 1702 AVPEHLCSWGPWPANWRC-----NITPC-----ENMECRDTRCYCEKVQLKLQOL 1748
Db 982 PAGRN-CEQMRPHSARCAVDEPLPPCMPTASALYQRDASQDSRFRCDIILKLFHSCDS 1040
QY 1749 SQFKSRCCGTC 1759
Db 1041 LEVRQKCCSTC 1051

RESULT 8
Q9EPX2 PRELIMINARY; PRT; 1280 AA.
AC Q9EPX2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PAPILIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1;
KC MEDLINE=20530499; PubMed=11076767;
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,
RA Steron A.L., Brockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the
RT ADAMTS metalloproteinases."
RL Development 127:5475-5485(2000).
DR EMBL: AF314171; AAG41980.1; -.
DR HSSP; P12111; 2KNT.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00047; Ig; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; IG_Like; 2.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS50092; TSPI; 3.
KW Immunoglobulin domain; Serine protease inhibitor.
SQ SEQUENCE 1280 AA; 138224 MW; AE287705E561AF30 CRC64;

```

Query Match

10.6%; Score 1019.5; DB 11; Length 1280;

Best Local Similarity 24.0%; Pred. No. 1.9e-65;

```

Matches 380; Conservative 153; Mismatches 527; Indels 523; Gaps 62;
QY 16 LAFLLSSRT---ARSEEDRDLWDAGPWSSECSFTCGGAGSYSLRCLSSK-----SC 66
Db 6 LFLSLLTSTPGSWANRRVQSDTWTGTGSEWSPCSSTCGGISFRERPCYSORRDGTS 65
QY 67 EGRNIRYRTCSNVDCPEAGDFRAQCSAHNDVKHHGQFVWLPVSNDDPNCSLKCOAK 126
Db 66 VGPARSRTCHTESCPDGYRDFRAEOCAFEDGTFDQGRYRWLPYYAAB-NKELNCIPK 124
QY 127 GTTLVVELAPKVDGTRCYTESLDMCISGLQCVGDHQLGSTVKEDNGCVNGDSTCR 186
Db 125 GONFYKHKDAVVDGTPCEPGQDICDGVGCRVWGVDHKLDSIKQEDKLCQGGDSSCY 184
QY 187 LVYRGKKSQLSKSDTVAIPYGSRHRLVLKGDPHLYLETKLQ-----TKG 237
Db 185 PVTGTFDGN-DISRGYNQIFIIIPAGATISRIEAAASRNFLAVKISIRGEYILNGHWTIEA 243
QY 238 ENSLSSTGTFLVDSNVDFQKFPDKEILRMAGPLTADETVKIRNSGSDSTVQIFYQPI 297
Db 244 AQALPVASTVLOYERCVEGDLAPER--IQARPTSEPIVELL-SQESNPGVHYEYLP 300
QY 298 I-----HWRETDFFPCSATCGGQYLTSAECYDLRSNRVADQYCHYYPENIKPKPKIQ 352
Db 301 NDPGRGFSWSHSGWGDCAECGGGHQSRVFC--TIDNEAYPDHMCQHQP---RPHR-R 354
QY 353 ECNLDPCCPASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRAVSCVEEDIQHV 412
Db 355 SCNTQPCPKT-----KRMKVGPWTPCVSCGGGVQSRYSYCISSDTGGO 399
QY 413 TSVEEWKCM-YTPKMPIAQPCNIFDCKPWLQAEWSPCTVTCGGQLRYRVVLCIDHRG--M 469
Db 400 EAAEETQCAGLAGKPTTQACNLQHCANVSEVPWCECVTCGTGINKRVTGCGDEGSP 459
QY 470 HTGCGSPYTKPHIKEECIVPTPCYKPKELPVEAKLPWFQAELEGAASVEEPSFTPE 529
Db 460 HAAACLLKQDPTLTPECV-----QEACPVFRGQAWH-----VG 492
QY 530 AWSACTVTCGVGTQVIRVRCQVLLSFSQSVADLPIDEC---EGPKPASORACYAGPCSGE 586
Db 493 SWSLCKSKCGGIRRRQVVCTI-----GPPRCVDTLQSSKPAEAMEACNRQPC--H 540
QY 587 IPEFNPDETLGLFGQLDFDELYDWEYEGYTKCE------SCGGGVQAEVASC 634
Db 541 LPQEVF-----SIQD-----PTRSSDPRLSGPRVSPVSDRGDQWAPLER 582
QY 635 LNKQT--REPAENLCVTSRRP-----POLLKS-----CNLDP---CFARW--EIG- 673
Db 583 PRAQSNPREGQDPNLSAGRAPTQRPHPQPLRPSGPRDCRHSPHGCCPDGHTFSLGP 642
QY 674 KWSPCSLTCGVGLQTRDVFCSHLIS-----REMNETVILLADELC 712
Db 643 QWQCPLAGASCLQSRYGCCPDGVSAEGPQOAGCTRSHGSDNTGNRPGSRVASKNPKI 702
QY 713 RQP-----KPSTVQACNRFNC-----PPAWYP----- 734
Db 703 HQQAHEGEPESECRS-SREGCCYDNVASAAGPLGBCGVQPSYAYPVRCILPSSAGSGCD 761
QY 735 --AQW-----QPCSRTCGG--GVQKREVLCORMA--D 761
Db 762 WAARWYFVASVGRNRFWYGGCHGNANFNASEQECMNTCRGHGPRRPEAGAAGRAHVD 821
QY 762 GSFLPELPETFCASAKPACQQAACKDDCPSEWLLSDWTECSTSCGEGTQTSATCRKMLKT 821
Db 822 G-----GORGPGQOE-----PDWHRAGATIPRLSPSGSPWRREQEP 859
QY 822 GLSTVWNSTLCPLPFPSSIRPCM--LATCARPGRSTKHSP-----HFAARKVYIOTRR 875
Db 860 APCEPPHIPAYGNRPGQEI RPRVPGDLREARPAVPT-HSPSYRIRLAGSEPSLQA-- 916
QY 876 QRKLFHVVGFGAYLLPKTAVVLRCPARVRKPLITWBDQOHLISSTHTVIVAPGYLKI 935
Db 917 -----APQAVQLFCPGNIPSEFQAGWQKEGRP-ISSNRYQLQADGSLIIS 961

```



Matches 371; Conservative 151; Mismatches 488; Indels 526; Gaps 65;

[illegible]

QY	942	AGVTC--SAGPAR--EHFVTKLIGGRKKLVARELSPRSEEVLGRKGGPKALQTHKHQ	998
Db	911	AGTSCGSTRPGRDSOKIQRITG---LCPHPIH-----HSH-944	
QY	999	NGIFSNGSKAEKRGLAANPSGRVDDILVLSRLLEGGWPGGELLASWFAQDSAEFNTTSEEDP	1058
Db	945	LVSPGLMTG---GDMAVLSEAE-----963	
QY	1059	GAEOVLLHLPTMTVEQRRLDDILGNLSOOPERLDRYLSKHLVAQLAQEIFRSHLEHODT	1118
Db	964	LSRFPQ-----969	
QY	1119	LLKFSERTSPVLSPHKHVGSGSSSLRTSSGTGAG--GGSRPRHXRKPTILRKISAAQOL	1176
Db	970	PRDPAQDFCQA-----GAAGPLGAIPSSHPQANRLRDQNO--1006	
QY	1177	SASEVWTHLGOTVALAGTSLVLLHCEAIGHPRPTISWNRNCEEVQFSDRILLQPDSDLQ	1236
Db	1007	PRVDASPGQIRWNTCRAGEFPFPAIEWORDGQPVSPRHLQLPDGSVLV1055	
QY	1237	ILAPVEADVGYTCNATFNALGYDS--VSIAY---TLACKPLVKTSMRTVNTIEKPAVTV	1290
Db	1056	ISRVAVEDGGFYTCVAFNGQDRQRVWQLRVLGELTISCLP-----PIVTV1101	
QY	1291	DIGSTIKTVGVNVTINCQVAGPEAEVTFWFRNKSGLSP-----HHLH-----EGSLLLTN1342	
Db	1102	PEGUTARLL-----CVVAG--ESYNIKWSRN---GLPVQADGHRVHQSPDGTLLIYN1148	
QY	1343	VSSDQGLYSACRANLHGLTESTQLLIIDPPQVPT1378	
Db	1149	LRADEGSYTCSAYQGSQVSRSTEYKVYSP--APT1182	

RESULT 11

076840 PRELIMINARY; PRT; 2167 AA.

ID 076840 Q22911;

AC 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE C37C3.6 PROTEIN.

GN C37C3.6.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Telodermata; Caenorhabditis.

OX NCBI\_Taxid=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Geisel C., Bradshaw H.;

RT "the sequence of C. elegans cosmid c37C3.";

CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY

CC CC -1- ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT

CC OF FORM B.

DR EMBL; U64857; AAC25868.1; -;

DR EMBL; U64857; AAC25867.1; -;

DR HSSP; P00981; IDTK.

DR InterPro; IPR003598; Ig\_c2.

DR InterPro; IPR003006; Ig\_MHC.

DR InterPro; IPR002223; Kunitz\_BPTI.

DR InterPro; IPR000884; TSP1.

DR Pfam; PF00047; Ig\_1.

DR Pfam; PF00014; Kunitz\_BPTI; 11.

DR Pfam; PF00090; tsp\_1; 6.

DR PRINTS; PR00759; BASICPTASE.

DR SMART; SM00408; IGC2; 1.

DR SMART; SM00131; KU; 11.

DR SMART; SM00209; TSP1; 7.

DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 10.

DR PROSITE; PS02079; BPTI\_KUNITZ\_2; 11.

DR PROSITE; PS50092; TSP1; 4.

[illegible]

```

QY 366 KQIMPYDLYHPLPWEATPWTACSSCG-GIQSRVAVSCVEDIQGHVTSVEENKCM--Y 422
Db 397 -----APHWEGEWSKSGKCGDGFQNRISITCRSSGHEVTVEEDAVCLKEV 445
QY 423 TPKMPIAPQCN--TFDCPKWLAQWSPCTVTCGGGLRVRVLCIDHRCMHHTGGSPKTKP 480
Db 446 GNKPATVQECNRDVKNCYKHLGPWTPCDKICGDKGQTKRKVTCFTIENGHK----- 496
QY 481 HIKBECIVPTPCYKPKKPLVPAKLPWFQKQALEGAASVEEPSFIPEANSACTVTCGV 540
Db 497 ----- 496
QY 541 GTQVRIVRQCVLLSFSQSADLPIDCEBGPAPQACRACYPAGCSETEPDPNDTDLFG 600
Db 497 -----LPDECEVEKPEKTEKSLLPCEG----- 522
QY 601 GLODFDELYDWEYEGFTKCSSECGGVQEAUVVCLNKTREPAENLCVTSRRPPQLLKS 660
Db 523 -----VDMILISWSSCN-ACQNTETRTAICGNKEGKYVPEE-FC--RPEVPTLSRP 570
QY 661 CNLDPCEARWEIGKWSLTCGVLQTRDVFCSHLLSREMNETVILADE--LC-RQPKP 717
Db 571 CKSPKCEAOWFSSEWSKCSAPCGKGVKSRIVICGEF-----DGKVTVPADDDSKCNKETKP 626
QY 718 STVQACNREN--CPANYPACWQPCSRFCGGGVQKREVLCQKRMADGSFLPEPFCAS 775
Db 627 ESEDCGEEKVCVGEWFTGPMGKSCPKCGGGERVREVLG---LSNGT-----KSVNCDDE 679
QY 776 K-PACQOACKKDDCPSEWLL 794
Db 680 KVEPLSEKCNSEACTEDEL 699

RESULT 13
Q9VAV4
ID Q9VAV4 PRELIMINARY; PRT; 3060 AA.
AC Q9VAV4; Q9VAV3;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE CG1540 PROTEIN.
GN PPN OR CG1540 OR CG18436.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa, Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler J., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC 1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AE003765; AAF56794.2; -
DR HSSP; P12111; 2KNT.
DR FlyBase; FBgn0003137; Ppn.
DR InterPro; IPR0003561; EGF-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR002221; WAP.
DR Pfam; PF00047; Ig_3.
DR Pfam; PF00014; Kunitz_BPTI; 12.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF00095; wap; 1.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00131; KU; 12.
DR SMART; SM00209; TSPI; 7.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS00317; 4-DISULFIDE CORE; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 11.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 12.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS00092; TSPI; 3.
KW Alternative splicing; Immunoglobulin domain;
KW Serine protease inhibitor.
FT VARSPPLIC 2803 2803 L-> SVVPV (IN SHORT ISOFORM).
FT VARSPPLIC 2844 2854 FNFKTMSDGI-> VASPLHPNAV (IN SHORT
FT VARSPPLIC 2855 3060 ISOFORM).
FT VARSPPLIC 3060 AA; 331579 MW; ACA31D3EE558C7C0 CRC64;
SQ SEQUENCE

Query Match 9.28; Score 882; DB 5; Length 3060;
Best Local Similarity 28.08; Pred. No. 8.6e-55;
Matches 224; Conservative 94; Mismatches 282; Indels 200; Gaps 29;

QY 37 DAWGPW---SECSRTCGGASYSILRRCL-----SSKSCGNRIYRVCNSVDCPEACDF 88
Db 18 DEWPWSPSDCSRTCGGVSYQTRCLRRDDRGEAVCSGSRIFSCNTODCEESDF 77
QY 89 RAQCSAHNDVKHGGQFYEWLPVSNDDPNPCSLKCAQGTLLVVELAPKVLDDGPRCYTES 148
Db 78 RAQCSRFDRQQDFGVFEWVPVYTNAP-NPCELNCMPKGERFYRQREKVDGTRCNDK 136
QY 149 LDMCISGLCIVGCDHQLGSTVVEDNCGVNGDSTCLRVGQYKSLSTKSDTVVAI 208
Db 137 LDVCVNECPVPCGDMMLGSDAKEDKCRKGGDGSTCKTIRNTITTKDLAPGYND-LLLL 195
QY 209 PYGSRHRLVLKGDHLYLETKLQG---TKGENSLSTGTFLVDNSVDVQK----FPD 261
Db 196 PGATNIRIETVPSNLYACRNHISGHYINGWRIDFPRMFANSWNYQRKPMGFAA 255
QY 262 KEILRMAGPLTAD-FIVKIRNSGSADSTVQIFVQPI-----IHRWRETDFFPCSATCG 314

```

Db 256 PDQLTSGPISESLFVLMVLQEKNIISLDYEYSIPESLSHSQDTHWTHQFNACSASCG 315  
QY 315 GGQVLTSAECYDLRSRV-----VADQYCHYYPENTKPKP-KLQECNLDPASPADGY 365  
Db 316 GGSQSRKVC-----NKRITLAEVNPISLQDQ-----KSKPVEQACGTEPC----- 356  
QY 366 KQIMPYDLYHPLPRWEATPWTACSSCGG-GIQSRAVSCVEEDIOGHVTSVEEWKCM--Y 422  
Db 357 -----APHWGEWSKSKSCGGSDGFQNRISITCERISSGHEHTEVEDAVCLKEV 405  
QY 423 TPAMPVIAQPN--IFDCKPKWLAQWSPCTVTCGGGLRYRVVLCIDHRCMTGGSPKTKP 480  
Db 406 GNPATKQECNRDVKNCPKYHLGFWPTPCDKLGDGKGQTRKVTCTFEBNGHK----- 456  
QY 481 HKEECIVPTPCYKPKELPVEAKLPWFQKQAELEGAASVEBPSFPIEWSACTVTCGV 540  
Db 457 ----- 456  
QY 541 GTQVRIVRQVLLSFSGVADLPIDECGPKPASQACQACQACQACQACQACQACQACQACQ 600  
Db 457 -----RV-----LPEDCVBEKPEKSKCLLTPCEG----- 482  
QY 601 GLQDFDELYDWEYEGFTKCSSESGGGVQAEVAVSCLNKQTPREAPENLCVTSRRPPQLLKS 660  
Db 483 -----VDWIIISQWSGCN-ACQONTETRTAICGNKEGVYBEE-FC--EPEVPTLSRP 530  
QY 661 CNLDPCEARWEIGKWSPCSLTCGVLQTRDVFCSHLLSRENNETVILADE--LC-ROPKP 717  
Db 531 CKSPKCEAQNFSSEWSKCAPCGKGVASRIVICGEF-----DGKIVTADDDSKCNKTEKP 586  
QY 718 STVQACNRFN--CPWAPYPAQWPCSTCGGQVQKREVLCKQRMADGSLFLELPTFCAS 775  
Db 587 ESEQDCGEKVCPCGWEFTGFWKCSKPCGGGERVREVLG---LSNCT-----KSVNCDEE 639  
QY 776 K-PACQACCKDDCPSEWLL 794  
Db 640 KVEPLSEKCNSEACTEDEL 659  
RESULT 14  
O44938 PRELIMINARY; PRT; 1572 AA.  
ID O44938;  
AC O44938;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE THROMBOSPONDIN.  
GN THRI.  
OS Haemochus contortus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Haemonchidae; Haemonchidae.  
OX NCBI\_TaxID=6289;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MOREDUN;  
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,  
RA Knox D.P.;  
RT "Cloning and characterization of thrombospondin, a novel multidomain  
glycoprotein associated with the gut of Haemonchus contortus.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043121; AAB99830.2; -;  
DR HSSP; P05067; ICA0.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00014; Kunitz\_BPTI; 12.  
DR PRINTS; PR00759; BASICPTASE.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00131; KU; 6.  
DR SMART; SM00209; TSP1; 7.  
DR PROSITE; PS02079; BPTI\_KUNITZ\_2; 6.  
DR PROSITE; PS00092; TSP1; 2.

KW Serine protease inhibitor.  
SQ SEQUENCE 1572 AA; 171871 MW; 2260B30DC2F903EC CRC64;  
Query Match 9.0%; Score 865; DB 5; Length 1572;  
Best local similarity 28.2%; Pred. No. 5.3e-54;  
Matches 224; Conservative 102; Mismatches 279; Indels 188; Gaps 30;  
QY 24 RTARSEEDRLDGLWDAWGPW--SECSRTCCGGGASYSLSRCLSSKSGEGRNIRYRCSNWD 81  
Db 66 KTGSES-----CPWGFVPEQCSRTCCGGVQTEKQCC--PGDCTGPSVYVSCNLEPC 117  
QY 82 PPEAGDFRAQOCSAHNDVKHGFYEWLPVSNPDNPSLCKQAKGTTLVVELAPKVLGD 141  
Db 118 -PECTDFAEQCAAHNDPDDGQYHKWIPYKG--KNKCELLCKPENGFFYKWDFTVVDG 174  
QY 142 TRCYTESLDACISLQIVGCDHQLGKSTIVKEDNGCVNGDGTCLRVGOY-KSOLSATK 200  
Db 175 TKDSKGGDDICVDGVCLPLGCDGKLGSALKLDKCGVCDGDSGQCKTEGSEDERNLSPGY 234  
QY 201 SDDTVVAIPYGSRRHRIWLVKGPDPHLYLETKTLOCTKGNSLSSTGTFLVDNSSVDFQK-- 258  
Db 235 HD--VMRIPAGATAIRIEEARPSSNNLAMKN-----SSDYIFLNGSMIQVEKD 282  
QY 259 -----FPD--KEILLRMAGPLTADFTIVKTR-NSGASDSTVQFIYQPII-----HRWRE 303  
Db 283 ELNGVHEFYDDGKPERITAKGPLLEEYVVSVLIRKGNRDSIKYEFVSPIYDEVDMYK 342  
QY 304 TDFPSCATCGGGYOLTSACYDLRSNHRVADQYCHYYPENIKPKKLOECNLDPASP 363  
Db 343 GEWSACSVCCKGQGTTPYCIETKTQARVDDQLC---DDANSTKPEFEK-----PCETVD 395  
QY 364 GYKQIMPYDLYHPLPRWEATPWTACSSCG--GGIQSRAVSCVEEDIOGHVTSVEEWKCM 422  
Db 396 QCAE-----WFKGEWEPSCQTCGQGEQYRVVYCHQVFANGRRITVDDGNC-- 441  
QY 423 TPKMP-IAQPCNIFDCPKWLAQWSPCTVTCGGGLRYRVVLCIDHRCMTGGCSPKTKPH 481  
Db 442 TAERPFAVRQVQCNRFACPEWQAGPWSACEKSGDAFYQKSVTC----- 483  
QY 482 IKECIVPTPCYKPKELPVEAKLPWFQKQAELEGAASVEBPSFPIEWSACTVTCGV 541  
Db 484 -----RSEKEGEGKL----- 494  
QY 542 TQVRIVRQVLLSFSGVADLPIDECGPKP-ASQACQACQACQACQACQACQACQACQACQ 600  
Db 495 -----LPAAECSSSENTLESQSCNLGPT----- 518  
QY 601 GLQDFDELYDWEYEGFTKCSSESGGGVQAEVAVSCLNKQTPRE-PAENLCVTSRRPPQLLK 659  
Db 519 GLKFFTS--EWKL-----CSK-CNDTEETREYTKDSQGRAYPLEKCLTDDEKEIPVDTR 570  
QY 660 SC-NLDPCEPARWEIGKWSPCSLTCGVLQTRDVFCSHLLSRENNETVILADELCRQPKPS 718  
Db 571 ACATQPCPNYEWTSQNSKSCSTECGHHKTRVICA---IHEGDVTVVDEALCOGEKPE 627  
QY 719 TVQAC-NRENCPPAPYPAQWPCSTCGGQVQKREVLCKQRMADGSLFLELPTFCASPK 777  
Db 628 VKMNCNTNEKCTGTWYSGPWSPCSGVECGGKQERVAVC-----LNYDKKFPVPEWCESEMP 683  
QY 778 ACQACKKDDCPES 790  
Db 684 NLIDPCNTDECPES 696  
RESULT 15  
Q908G8 PRELIMINARY; PRT; 3198 AA.  
ID Q908G8;  
AC Q908G8;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE LACUNIN PRECURSOR.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 02:20:32 ; Search time 61.07 Seconds  
(without alignments)  
3204.716 Million cell updates/sec

Title: US-10-044-807-2  
Perfect score: 9588  
Sequence: 1 MECCRRATPGTLLFLAFLL.....LKLCQLSFKSCCGTCGKA 1762

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9466.5	98.7	1745	22 AAB90551	Human secreted pro
2	9464.5	98.7	1766	22 ABB10246	Human CDNA SEQ ID
3	4797.5	50.0	874	21 AAB01431	Human TANGO 224 (f
4	3956.5	41.3	1691	22 AAE07863	Novel human protei
5	3953.5	41.2	1691	22 AAE10609	Human novel KIAA12
6	3762.5	39.2	1617	22 AAE07870	Novel human protei
7	3565	37.2	643	21 AAY87347	Human signal pepti
8	3423	35.7	645	22 ABB10454	Human CDNA SEQ ID
9	2917	30.4	525	20 AAY41729	Human PRO1071 prot
10	2917	30.4	525	22 AAU29241	Human PRO1071 UNQ
11	2907	30.3	525	21 AAB44285	Human PRO1071 UNQ

12	2907	30.3	525	22 AAB72290	Human ADAMTS-R1 am
13	2809.5	29.3	845	22 AAE07868	Novel human protei
14	2615.5	27.3	771	22 AAE07869	Novel human protei
15	2588	27.0	514	22 AAB90586	Human secreted pro
16	2567.5	26.8	1281	22 AAE10608	Human novel KIAA12
17	2567.5	26.8	1281	22 AAE10610	Human novel KIAA12
18	2425.5	25.3	724	22 AAE07866	Novel human protei
19	2283.5	23.8	417	22 AAG78679	Human thrombotic p
20	2282.5	23.8	926	22 AAB83370	NOV14 protein sequ
21	2231.5	23.3	650	22 AAE07867	Novel human protei
22	2198	22.9	428	21 AAB41561	Human ORFX ORF1325
23	1981	20.7	480	21 AAB01425	Human TANGO 224.
24	1930	20.1	368	22 ABB11218	Human BPTI/Kunitz
25	1505	15.7	292	22 AAB90632	Human secreted pro
26	1438	15.0	446	22 AAE07864	Novel human protei
27	1388	14.5	262	22 AAB90587	Human secreted pro
28	1353	14.1	886	21 AAB40500	Human secreted pro
29	1244	13.0	372	22 AAE07865	Human ORFX ORF264
30	1145.5	11.9	2150	21 AAY53898	Novel human protei
31	1145.5	11.9	2165	22 AAB90617	Amino acid sequenc
32	1111	11.6	281	22 ABB11315	Human secreted pro
33	985.5	10.3	1934	22 AAB72301	Human HSP-124 pro
34	972.5	10.1	1862	22 AAB72286	Human ADAMTS-9 alt
35	961	10.0	180	20 AAY35899	Extended human sec
36	947	9.9	337	22 ABG27865	Novel human diagno
37	886	9.2	207	22 AAM25476	Human protein sequ
38	884.5	9.2	1784	21 ABA11379	Human ORFX ORF1143
39	882	9.2	2858	22 ABB71150	Drosophila melanog
40	882	9.2	3060	22 ABB58084	Drosophila melanog
41	870	9.1	169	22 ABG27864	Novel human diagno
42	843	8.8	1686	22 AAB74944	Human ADAM type me
43	831	8.7	1686	22 AAE00913	Human 27875 ADAM-T
44	831	8.7	1690	22 AAB86949	Human metalloprote
45	830	8.7	1686	22 AAE00934	Human 27875 ADAM-T

ALIGNMENTS

RESULT 1  
AAB90551  
ID AAB90551 standard; Protein; 1745 AA.  
XX AC AAB90551;  
XX DT 01-JUN-2001 (first entry)  
XX DE Human secreted protein, SEQ ID NO: 89.  
XX DE Human; secreted protein; immunomodulatory; antisclerotic;  
XX KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;  
XX KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
XX KW neurotropic; anticonvulsant; antialzheimers; antiparkinsonian;  
XX KW antimicrobial; vulnerary; vaccine; gene therapy; cancer;  
XX KW protein coordinate data; infection.  
XX OS Homo sapiens.  
XX PN WO200121658-A1.  
XX PD 29-MAR-2001.  
XX PF 22-SEP-2000; 2000WO-US26013.  
XX PR 24-SEP-1999; 99US-0155709.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Ni J. Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
XX PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
XX PI Young PE, Wei P, Florence KA;  
XX DR WPI; 2001-235311/24.

*Requested Here*

DR	N-PSDB; AAF97891.	601	GLQDFDELYDWEYEGFTKCSGCGGVOEAVVSCLNKOTREPAENLVCVTRRRPQLLKS	660
XX	Nucleic acids encoding 32 human secreted polypeptides, useful for			
PT	preventing, diagnosing and/or treating e.g. cancers, Parkinson's	584	glqdfdelwyeyegftkcsescgvgqavvsclnkqtrepaeenlcvtstirrpqllks	643
PT	disease and diabetic retinopathy -			
XX	Claim 11; Fig 4; 890pp; English.	661	CNLDPCPARWEIGKWSPCSLTCGVLQTRDVFCSHLLSRENNETVILADELCROPKSTV	720
PS				
XX	The present sequence is one of 32 novel human secreted polypeptides. The	644	cnldpcparweigkwspscltcgvgltqtrdvfcshllsremnetvilladelcrpkipstv	703
PS	nucleic acid molecules and polypeptides may be used in the prevention,			
CC	diagnosis and treatment of diseases such as immune disorders (e.g.	721	QACNRENCPPAWYPAQWOPCSTRCTGGGVQKREVLCKQRMADSGSELELPETFCSSKAPACQ	780
CC	multiple sclerosis, systemic lupus erythematosus and human			
CC	immuno-deficiency virus (HIV) infections), hyperproliferative disorders	704	qacnrcfcpawypaqwpcstrctggvgqkrevlckqrmadsgflepeticssaskpacq	763
CC	(e.g. cancers and Gaucher's disease), cardiovascular diseases			
CC	(e.g. Sclimtar syndrome, Chaga's cardiomyopathy and coronary	781	QACKKDDCPSWLLSDWTFECTSCGEGQTRSAICRMLKMTGLSTVNVNSTICPPLPSSS	840
CC	arteriosclerosis), angiogenic disorders (e.g. corneal graft			
CC	neovascularisation and diabetic retinopathy), neurological disorders	764	qackkddcspewllsdwtectscgegqtrsaicrmlkmtglstvnvusticpplpfess	823
CC	(e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),			
CC	infectious diseases and/or for promoting wound healing, regeneration	841	IRPCLATCAPRSPSTKHSPIAARVYIOTRQRLHFVGVGFAYLLPDKTAVVLRCP	900
CC	and/or chemotaxis. The nucleic acid molecules may be used to produce the			
CC	secreted polypeptides. They may also be used as DNA probes in diagnostic	824	irpclatcarprpstkhspiaaarvylqtrrrqkhrfvvggfayllpdktavvlrcp	883
CC	assays to detect and quantitate the presence of similar nucleic acid			
CC	sequences in samples. The polypeptides may be used as antigens in the	901	ARRVRKPLITWEKDGQHLISSTHTVAPFGYLKIHRLKPSDAGVYTCISAGPAREHFVKL	960
CC	production of antibodies and in assays to identify modulators of			
CC	their expression and activity.	884	arrvrkplitwekdqghlissthtvafgykhrlkpsdagvytcsagparehfvkl	943
XX	Sequence 1745 AA;			
SQ		961	IGNRKLVARPLSPRSEEVLAGRGGPKALQTHKHONGIFSNCSKAERKGLAANPGR	1020
		944	iggnrkliavarplsprseevlagrkgpkalqthkhqngifsnsgskaerkglaanpgr	1003
		1021	YDDLVRLLLEOGWPGGELLASWEAODSAERNTTSEEDPGAEOVLHLPFTWTEORRLDD	1080
		1004	yddlvrlllegggpggellasseaqsarnttseedpgaevllhlpftwteqrldd	1063
		1081	ILGNSLQOPEELRDLYSKHLVAQAEIFRSHLEHQDTLLKPSERTSFVTLSPHKKHVS	1140
		1064	ilgnlsqpeelrldylskhlvaqlaeqfirshlehdqtlkpsertspvtilsphkhvsg	1123
		1141	FSSLSRSTGDGCGSRPRHKKPTILRKISAAQOLSAFVVTHLQGVVALASGTLVLL	1200
		1124	fsslsrstgdaggsrrphrkptilrkisaaqqlsasevthlqgvvalasgtlsvll	1183
		1201	HCEAIGHPRPTISARNGEVQFSDRIQLQDDSLQIILAPVEADVFTCNATNALGYDS	1260
		1184	hceaighprptisarngeevqfsdrillqpdslqilapveadvfytchnatnalgysd	1243
		1261	VSIATVLAGPLVKTISRTMINTKPAVTVDIGSTIKTVQGVNVTINCOVAGVPEAEVTW	1320
		1244	vsiavtlagkplvktisrmtintekpavtdigstiktvggvnvtincqvagvpeaevtw	1303
		1321	FRNKSGLSPHLLHSGSLLLTNVSSDGLYSCRAANLHGETESTQILLILDPPQVPTQL	1380
		1304	frnkskigsphllhsgsllltnvssddgylsraanlhgeltetqllildppqvptql	1363
		1381	EDIRALLAATGPNLPSVLTSPGLTOLVDGPNNSALLGCPKIGHPPVNTWTFHGGPIVTA	1440
		1364	edirallaatgpnlpstvltspgltdlvdgpnnsallgcpkighppvntwtfhggpipta	1423
		1441	TGLTHHILAAQGILOVANLSGGSGQFESCLAQNEAGVILMQKASLVIOQYWNVSVDLATCS	1500
		1424	tglthhilaagqilqvanlsggsgqfescslaagneagvimqkaslvioqywnvsdlatcs	1483
		1501	ASCGNRGVQOPLRLCLINSTEVPNAHCAGKVPAPQIACNRRDCPSRMWTSWACVTR	1560
		1484	ascgnrgvqqplrlcllnstevnpahcagkvpavqplacarrdpcpsrmwtswactirs	1543
		1561	CGGVQTRRVTCQKLKASGISTPVSNDMCTOVAKRPVDTQACNQQLCVWEAFSSWGQNG	1620
		1544	cggvqtrrvtcqklkasgistpvsndmctqvkrpvdtaqcnqlcvewafsswgqng	1603
		1621	PCIGPHLAVQHROVFCQTRDGTITLPEQCSALPRPVSTQNCWSEACSVHWRVSLTCLTA	1680
		1604	pcigphlavqhrovcqtrdgtitlpeqcsalprpvstqncwseacsvhwrvsletclta	1663
		1681	TCGNVFGFSRRVECVHARTNKAVPEHLCSWGPANQRCNITPCENMECRDTRTYCEKV	1740

Query Match 98.7%; Score 9466.5; DB 22; Length 1745;  
 Best Local Similarity 99.0%; Pred. No. 0;  
 Matches 1744; Conservative 1; Mismatches 0; Indels 17; Gaps 1;

QY	1	MECCRRATPGTLILFLALLSSRTARSEEDRGLDWAQWPWSCSTRCTGGGASYSLRRC	60
DB	1	mecorrtatpgtllflallssrtarseedrdglwawpwwscstrctgggasyllrrc	60
QY	61	LSSKSCBGRNIRYTCNVDCPPBAGDRAQQCSAHNDVXHHGQFYEWLPSVNDPNDPCS	120
DB	61	lsskscegnlryrtcsnvdcppeagdraqqcsahndvkhgqfyewlpvsndpndpcs	120
QY	121	LKCAQKCTVLVAPLAKVLDGTRCYTESLDMCLISGLCQIVGCCDHOIGSTVKEDNCGVNG	180
DB	121	lkcaqkctvlvaplakvldgtrcytesldmclisglcqvigcdhqlgstvkednvcvng	180
QY	181	DGSTRCLVRGQYKSQLSATKSDDPVVAIPYXGSHRILVKGPDHLYLETFLQGTGKENS	240
DB	181	dgstrclvrqyksqlsatksddtvaipysrhrilvkgpdhlyletklqgtkngens	240
QY	241	LSSTGTFELVONSVDFOKFPDKKEILRMAGPLTADFTVIRKNSGSADSTVQPIFYQPIIHR	300
DB	241	lsstgtflvdnssvdfkfpdkelrmagpltdftvirknsgsadstvgfifyqpiih	300
QY	301	WRETDFFPCSATCGGQYOLTSAECDYDLRSNRVVADQYCHYYPENIKPKPLQECNLDPCP	360
DB	301	wrecdfpcsatcggyqltsaecdydlrsnrvadqychyypenikpkplqecnldpcp	360
QY	361	ASDGYKQIMPYDLVHPLFRWEATPWTACSSSCGGGIGQSRVSCVEEDIQGHVTSVEEWKC	420
DB	361	a-----rweatpwtacssscggigqsravscveedighvtsveewkc	403
QY	421	MYTPKMPIAQPCNITFDCKPKIAQWSPCTVTCGGGLRYRVVLCIDHRCMHVTCGSPKTKP	480
DB	404	mytpkmpiaqpcnifdcpkwlqagwspctvtcggglyrvvvlcidhrgmhtggcspktp	463
QY	481	HIKEECIVPTCYKPKELPVEAKLPWFQAOLEBGAASVEEPSTFPEAWSACTVTCGV	540
DB	464	hikeecivptcykpkelpreaklpwfkaqlaeeagaasveepsfipkawsactvtcv	523
QY	541	GTQVRIVRQCVLLSFSQVADLPIDECEGPKPASQRACYPGCSGEIPEFNPDETDGLFG	600
DB	524	gtqvrivrqcvllsfsqsvadlpidecegpksqracyagpcsgiepefnpdetdglfg	583

|||||  
Db 1664 tcgnygfgrvccvhartnkavpehlcswgprpanwgrcnitpcenmecrtdtrycekv 1723  
QY 1741 KQLKLCQLSQFKSRCCGCGKA 1762  
Db 1724 kqkicqisqfskrcgcgcka 1745  
RESULT 2  
ABBI0246  
ID ABB10246 standard; Protein; 1766 AA.  
XX ABB10246;  
XX  
DT 10-JAN-2002 (first entry)  
XX  
DE Human cDNA SEQ ID NO: 554.  
XX  
KW Human; gene therapy; neural disorder; immune system disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
KW proliferative disorder; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200154474-A2. *Aggregated have*  
XX  
PD 02-AUG-2001.  
PF 17-JAN-2001; 2001WO-US01349.  
XX  
PR 31-JAN-2000; 2000US-179065P.  
PR 04-FEB-2000; 2000US-180628P.  
PR 24-FEB-2000; 2000US-184664P.  
PR 02-MAR-2000; 2000US-186350P.  
PR 16-MAR-2000; 2000US-189874P.  
PR 17-MAR-2000; 2000US-190076P.  
PR 18-APR-2000; 2000US-198123P.  
PR 19-MAY-2000; 2000US-205515P.  
PR 07-JUN-2000; 2000US-209467P.  
PR 28-JUN-2000; 2000US-214886P.  
PR 30-JUN-2000; 2000US-215135P.  
PR 07-JUL-2000; 2000US-216647P.  
PR 07-JUL-2000; 2000US-216880P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217496P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.  
PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225213P.  
PR 14-AUG-2000; 2000US-225214P.  
PR 14-AUG-2000; 2000US-225266P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 14-AUG-2000; 2000US-225758P.  
PR 14-AUG-2000; 2000US-225759P.  
PR 18-AUG-2000; 2000US-226279P.  
PR 22-AUG-2000; 2000US-226681P.  
PR 22-AUG-2000; 2000US-226868P.  
PR 23-AUG-2000; 2000US-227182P.  
PR 23-AUG-2000; 2000US-227009P.  
PR 30-AUG-2000; 2000US-228924P.  
PR 01-SEP-2000; 2000US-229287P.  
PR 01-SEP-2000; 2000US-229343P.  
PR 01-SEP-2000; 2000US-229344P.  
PR 01-SEP-2000; 2000US-229345P.  
PR 05-SEP-2000; 2000US-229509P.  
PR 05-SEP-2000; 2000US-229513P.  
PR 06-SEP-2000; 2000US-230437P.  
PR 06-SEP-2000; 2000US-230438P.  
PR 08-SEP-2000; 2000US-231242P.  
PR 08-SEP-2000; 2000US-231243P.  
PR 08-SEP-2000; 2000US-231244P.  
PR 08-SEP-2000; 2000US-231413P.  
PR 08-SEP-2000; 2000US-231414P.  
PR 08-SEP-2000; 2000US-232080P.  
PR 08-SEP-2000; 2000US-232081P.  
PR 12-SEP-2000; 2000US-231968P.  
PR 14-SEP-2000; 2000US-232397P.  
PR 14-SEP-2000; 2000US-232398P.  
PR 14-SEP-2000; 2000US-232399P.  
PR 14-SEP-2000; 2000US-232400P.  
PR 14-SEP-2000; 2000US-232401P.  
PR 14-SEP-2000; 2000US-233063P.  
PR 14-SEP-2000; 2000US-233064P.  
PR 14-SEP-2000; 2000US-233065P.  
PR 21-SEP-2000; 2000US-234223P.  
PR 21-SEP-2000; 2000US-234274P.  
PR 25-SEP-2000; 2000US-234997P.  
PR 25-SEP-2000; 2000US-234998P.  
PR 26-SEP-2000; 2000US-235484P.  
PR 27-SEP-2000; 2000US-235834P.  
PR 27-SEP-2000; 2000US-235836P.  
PR 29-SEP-2000; 2000US-236327P.  
PR 29-SEP-2000; 2000US-236367P.  
PR 29-SEP-2000; 2000US-236368P.  
PR 29-SEP-2000; 2000US-236369P.  
PR 29-SEP-2000; 2000US-236370P.  
PR 02-OCT-2000; 2000US-236802P.  
PR 02-OCT-2000; 2000US-237037P.  
PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 02-OCT-2000; 2000US-237040P.  
PR 13-OCT-2000; 2000US-239935P.  
PR 13-OCT-2000; 2000US-239937P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241221P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241786P.  
PR 20-OCT-2000; 2000US-241787P.  
PR 20-OCT-2000; 2000US-241808P.  
PR 20-OCT-2000; 2000US-241809P.  
PR 20-OCT-2000; 2000US-241826P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 08-NOV-2000; 2000US-246474P.  
PR 08-NOV-2000; 2000US-246475P.  
PR 08-NOV-2000; 2000US-246476P.  
PR 08-NOV-2000; 2000US-246477P.  
PR 08-NOV-2000; 2000US-246478P.  
PR 08-NOV-2000; 2000US-246523P.  
PR 08-NOV-2000; 2000US-246524P.  
PR 08-NOV-2000; 2000US-246525P.  
PR 08-NOV-2000; 2000US-246526P.  
PR 08-NOV-2000; 2000US-246527P.  
PR 08-NOV-2000; 2000US-246528P.  
PR 08-NOV-2000; 2000US-246532P.  
PR 08-NOV-2000; 2000US-246609P.  
PR 08-NOV-2000; 2000US-246610P.  
PR 08-NOV-2000; 2000US-246611P.  
PR 08-NOV-2000; 2000US-246613P.  
PR 17-NOV-2000; 2000US-249207P.  
PR 17-NOV-2000; 2000US-249208P.  
PR 17-NOV-2000; 2000US-249209P.  
PR 17-NOV-2000; 2000US-249210P.  
PR 17-NOV-2000; 2000US-249211P.  
PR 17-NOV-2000; 2000US-249212P.  
PR 17-NOV-2000; 2000US-249213P.  
PR 17-NOV-2000; 2000US-249214P.  
PR 17-NOV-2000; 2000US-249215P.  
PR 17-NOV-2000; 2000US-249216P.  
PR 17-NOV-2000; 2000US-249217P.



Db 1445 tglthhlaagqilqvanlsgsgqsfclaqneagvlmqkaslvigdywsvdrlates 1504  
QY 1501 ASCNGVQOPRLRCLLNSDEVNPAHCAGKVRPAVOP IACNRRDCPSRMVMTSWSACTRS 1560  
Db 1505 ascngvqvprrcllnstevnpahcagkvrpavpapiacnrrdcpsrmvmtswsactrs 1564  
QY 1561 CGGGVTRRTVTCOKLKASGISTPVSNDMCTQVAKRPVDTQACNQQLCWEAFSSWGQCNG 1620  
Db 1565 cggvgvtrrvtcqklkasglistpvsndmctqvakrpvdtqacnqqlcweafsswgqng 1624  
QY 1621 PCIGPHLAVQHRVFCQTRDGIITLPSEQCSALPRPVSTQNCWSEACSVHVRVSLWLICTA 1680  
Db 1625 pcigphlavghrqvfqctrdgitlpseqcsalprpvstqncwseacsvhvrslwlclcta 1684  
QY 1681 TCQNYGFSQSRVRVCHVHARTNKAVPEHLCSMGPRPANQRCNITPCCNMECRDTRYCEKV 1740  
Db 1685 tcqnygfsqrvecvhartnkavpehlcsmgprpanqrcnltcpnemecrdttrycekv 1744  
QY 1741 KQIKZQLSQFQRSRCCTGCKA 1762  
Db 1745 kqiklqslsqfksrccgtcgka 1766

RESULT 3

AAB01431  
ID AAB01431 standard; Protein: 874 AA.

XX AAB01431;

DT 20-OCT-2000 (first entry)

XX Human TANGO 224 (form 2).

XX TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
KW inflammatory bowel disease; septic shock; ulcerative colitis;  
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
KW systemic lupus erythematosus; transgenic animal; diagnosis;  
KW prognosis; prophylactic; therapeutic; human.

XX Homo sapiens.

XX WO200039284-A1.

XX 06-JUL-2000.

XX 23-DEC-1999; 99WO-US31025.

XX 30-DEC-1998; 98US-0223546.

XX (MILL-) MILLENNIUM PHARM INC.

XX Holtzman DA;

XX WPI; 2000-465743/40.

XX N-PSDB; AAA47482.

XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases

XX Claim 8; Fig 30; 209pp; English.

XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
CC agents for regulating cellular processes like asthma, graft  
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
CC lupus erythematosus. The nucleic acids are also useful for producing

CC transgenic animals and the TANGO polypeptides themselves. Partial  
CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
CC forensic biology, for diagnostic assays, prognostic assays,  
CC pharmacogenomics and for monitoring clinical trials. TANGO  
CC polypeptides are suitable for both prophylactic and therapeutic  
CC methods for treating a subject at risk of a disorder or having a  
CC disorder associated with aberrant TANGO expression. A wide range  
CC of cellular disorders can be treated.  
XX  
SQ Sequence 874 AA;  
  
Query Match 50.0%; Score 4797.5; DB 21; Length 874;  
Best Local Similarity 98.9%; Pred. No. 2 3e-315;  
Matches 864; Conservative 1; Mismatches 0; Indels 9; Gaps 1;  
  
QY 1 MECRRATPTGTLTLFLAFLLLSSWTARSEEDRGLDAMGFWPSCSRTOGGGASYSILRR 60  
Db 1 mecrratptgtlllflaflllssrtarseedrdglawdgpwsecsrctcgggasyilrrc 60  
QY 61 LSSKSCGRNLRVTCNVDCPPPEAGDFRAQOCSAHNDVKHGOFYEWLPVSNDDNPCS 120  
Db 61 lsskscegrnirvrtcsnvdcpppeagdfraqqcsahndvkhgqfyewlpvsnddnpcs 120  
QY 121 LKQAKGTTLLVELAPKVLDTGTCYTESLDMCISGLCQIVGCDHQLGSTVKEDNCGVNG 180  
Db 121 lkqakgttllvelapkvldgttrcytesldmcisglcqiivgcdhqlgstvkedncgvng 180  
QY 181 DGSTCRIVRGQYKSQLSATKSDDTTVAIPYGSRHRIVLKGPDBHLYLETTLQGTKGENS 240  
Db 181 dgstcrivrvgqyksqlsatksddttvaipygsrhirvlkghpdlhlylettlqgtkgens 240  
QY 241 LSSTGTLVDSNVDFDKFDPKKEILRMAGPLTADFIKIRNSGSDSTVOFIFQPIIHR 300  
Db 241 lssutgtlvdsnvdfdkfdpkellrmagpltdafikirnsdsadstvffiqpilihr 300  
QY 301 WRETDFPCSATCGGGYQLTSAECYDLRSNRVADQYCHYYPENIKPKLQECNLDPCP 360  
Db 301 wretdfpcsatcggyqltsaecydlrsnrvadqychyyypenikpkqlqecnldpcp 360  
QY 361 A-----SDGKYQIMPYDLYHPLPRWEATPWATGSSCGGGIQSRANVSCVEEDTOGH 411  
Db 361 aqrvgiivpnsgdykqimpydlyhplprweatpwatgsscgggigravscveediqgh 420  
QY 412 VTSVEWKMCTPKMPTAOPCNIFDCPKWLAQWSPCTVTCGGRLRYRVILCIDHRGHT 471  
Db 421 vtsveewkcmtytpkmpiaqpcnifdcpkwlaqewspctvtcgggryrvilcidhrghmt 480  
QY 472 GGCSPKTKPHKEECIVPTFCYKPKELPVEAKLPWFKAQELEEGAAVSEEPSFIFEAW 531  
Db 481 ggcspktkphikeecivptfcykpkelpveaklpwfkaqelegeaavseepsfipeaw 540  
QY 532 SACTVTGCVGTQVRIVRCQVLLSFSQSVDLPIIDBCGPKPASORACYPAGPCSEIPEFN 591  
Db 541 sactvtcgvgtvrivrcqvllsfsqsvadlpidecegpkpasgracyagpcseipefn 600  
QY 592 PDETDGLFGGLQDFDELYDWEYEGFTKCSGSGGVQEAQVAVVCLNKOTREPAENLCVTS 651  
Db 601 pdetdglfgglqdfdelydweyegftkcsescgvgvqeaavvclnktrepaenlcvts 660  
QY 652 RRPQLKSNLDPCPARWEIGKWSPCSLTCGVGLQTRDVFCSHLLSRENETVILADEL 711  
Db 661 rrpqllksnldpcparweigkwspscltcgvglqtrdvfcshllsrennetviladel 720  
QY 712 CRQPKSTVQACNRFNCPNPPAWYPAQWQPCSRTOGGGVQKREVLCQKRMAGSGFLELPETF 771  
Db 721 crqpkstvtqacnrfncppnppawypaqwpcsrtooggvqkrevlckqrmagsgflepETF 780  
QY 772 CSASKPACQACQKDDCPSFWLLSDWTECSTSCGEGTQTRSAICRKMILKTGLSTVNVSTL 831  
Db 781 csaskpacqacqkddcpsfwllsdwtecsstscgegtqtrsaicrkmilktglstvnvstl 840  
QY 832 CPPLPFSSSIRPCMLATCARPGRPSTKHSPHIAA 865





```
QY 1311 AGVPAEAVTWFRNKSGL-GSPHLLHEGSLLTNNVSSDOGLYSCRAANLHGELTESTQLL 1369
Db 1324 kgypnptwkrqsgslgnvslfngslilqnvslenegtyvciatnalgkavatsvhh 1383
QY 1370 IIDPPQVPTQIEDIRA----LLAATGPNLPVSLTSPGLTQLVLDPNSALLGCPYKHPV 1425
Db 1384 llerwpesrivilgghkkyilqatnrtns-----ndp-----tgepp 1422
QY 1426 PNITWFGGQPIVTATGLTHHILAAQIILQVANLSGSGSGEFSCLAQNEAGVLMQKASIV 1485
Db 1423 p----- 1423
QY 1486 IQDYWMSVDRLATCSASCGNRGVQOPRLRLCLL-NSTEVPNPAHCAGKVRPAVQPIACNRRD 1544
Db 1424 -qepfwepgnwshscatcghlqariqpqcmvangevsealcldhqlkplagfepcnird 1482
QY 1545 CFSRWMTVSWACTSCGGGVTRRTVCOKLKASGISTVPSNDMCTQVAKRPVDTQACNQ 1604
Db 1483 cparwftsvsqscvscgegyhsrqvctkrkangltvqvvspracap-kdrplgrkpcfg 1541
QY 1605 QLCVEMAFSSWQCNCPCIGPHLAVQHROVFCQTRDGIPLPSEQCSCALPRPVSTQNCWSE 1664
Db 1542 hpcvqw--epgnrcpgrcmgravrqmqrhtacqhns-----dsncodrkrptlrrnctsg 1595
QY 1665 ACSVHWRSWLTLCTATCGNYGFSRRVCEVHARTNKAVPEHLCSWGPRPANWQRCNITP 1724
Db 1596 acdvcwhitgpkpctaacgr-gfgrskvdcitrsckpvaktrchvqkxkpwiswhc-lgp 1653
QY 1725 CENMECROTTRICEKVQKQLQSLQSFKRCGTC 1759
Db 1654 scdrctdtdthycmfvkhnlnclsldrykrcqsc 1688

RESULT 5
AAE10609
ID AAE10609 standard: Protein; 1691 AA.
XX AC AAE10609;
XX DT 10-DEC-2001 (first entry)
XX DE Human novel KIAA1233-like protein, NOV-2a.
XX KW Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
KW cytotstatic; nootropic; anti-infertility; cancer; chromosome 15;
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Active-site 1..26
FT /label= Signal_peptide
FT Protein 27..1691
FT /note= "Human mature novel KIAA1233-like protein,
FT NOV-2a"
XX WO200162928-A2.
XX 30-AUG-2001.
XX 26-FEB-2001; 2001WO-US06151.
XX 25-FEB-2000; 2000US-0184951.
XX 28-FEB-2000; 2000US-0185548.
XX 01-MAR-2000; 2000US-0185967.
XX 18-APR-2000; 2000US-0197723.
XX 27-APR-2000; 2000US-0199957.
XX 23-FEB-2001; 2001US-0789390.
```

```
XX PA (CURA-) CURAGEN CORP.
XX PI Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
XX WPI; 2001-582051/65.
XX N-PSDB; AAD17757.
XX PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
XX PT polypeptide for diagnosing and treating pathological disorders, such as
XX PT Parkinson's disease and for use in pharmacogenomics -
XX PS Claim 1; Page 22-23; 189pp; English.
XX The invention relates to novel human polypeptides referred as NOV-X
XX and their corresponding nucleic acid sequences. NOV-X collectively
XX include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
XX polypeptides, NOV-3b, NOV-3c and NOV-3d which are novel
XX STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
XX which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
XX identify a potential therapeutic agent that can modulate its activity
XX and can be used for treating a pathology related to aberrant expression
XX or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
XX used to determine the presence or predisposition to a disease associated
XX with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
XX to treat or prevent a pathology associated with NOV-X. The pathological
XX states that can be treated or prevented are haematopoietic, cancer,
XX immunological, tumour, neurodegenerative (e.g. Alzheimer's and
XX Parkinson's disease), human immunodeficiency virus (HIV) illness and
XX fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
XX predictive medicine. NOV-X DNA is used in gene therapy. The present
XX sequence is human novel KIAA1233-like protein, NOV-2a. NOV-2a gene is
XX located at chromosome 15.
XX SQ Sequence 1691 AA;
```

Query Match 41.2%; Score 3953.5; DB 22; Length 1691;

Best Local Similarity 43.3%; Pred. No. 7.2e-258;

Matches 768; Conservative 265; Mismatches 555; Indels 187; Gaps 31;

```
QY 22 SSTARSEDRDGLDAMGPMWSECSRTCCGGASYSRLRCLSSKSCGRNIRYTCNSVDC 81
Db 64 tsnrtsrdsdcdgndawgdscrtccggasysrlrcltgrncegnnrytkcsnhdc 123
QY 82 PPBAGDFRACQCAHNDVKKHGFYEWLPVSNPDNPNCSLKCOAKGTTIVVELAPKVLG 141
Db 124 pdaeifraqcsayndvqgghyewlpryndpaapcalckhaqgnlvvelapkvldg 183
QY 142 TRCYTESLDMCISGLCQIVGCDHQLAGSTVAKEDNCGVCNGDGTCLRLVGRQYKSQLSATKS 201
Db 184 trcntdsldmcisgicqavgcdrlqgsnakedncgvcagdgstclrvrgskshvapekr 243
QY 202 DDTWVAIPGYSRHIRLVKLPDHLXLETKTLQGTGKNSLSSTGTFLVDNSSVDFOKFPD 261
Db 244 eenviavpigrsvritvkgpahlfiesktlqgskgshsnspgvfvventtvefgqse 303
QY 262 KEILRMAGPLTADIVKIRNSGSDSTVQIFVQPIIHRWRRETDFFPCSATCGGQULTS 321
Db 304 rqtfkfpgplmadfifktrytaaksvvqffqgishqwrqtdffpctvtcgggyqlns 363
QY 322 AECYDLRSNRVADQYCHYYPENIKPKQLQECNLDPCPASDGYKOLMPYDLYHPLPRWE 381
Db 364 aecvdrlkrvpdhychyyppenvkpkklkecsmdpcssddfkemlypdyhfpqprwe 423
QY 382 ATPWTACSSSCGGGIQSRVAVSCVEEDIQGHVTSVEEWKCMYTPKMPAQPCNIFDCPKWL 441
Db 424 hnpwtacsvscgggiqrsvfvcveesmhgeilqveewkcmypkpmqtcnlfidcpkwi 483
QY 442 AQEWSPTCTVTCGGGLRYRVVLCIDHRGMHTGCGSKPKPHKEECIVPTPCYKPKELPV 501
Db 484 amewsgctvtcgrglrvrvvclnhrghvggcnpgnklhkeecvpiipcykpkexspv 543
```



Db 32 rncegqnrkykcsnhdcpdaedfraagcsayndvqyghyewlpryndaapcalkc 91  
QY 124 QAKGTTLVLAPKVLDTGRTYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNGDGS 183  
Db 92 haqgnlvvelapkvldgtrcndtsldmcisgicqavgcdrlqgsnakedncvcagdg 151  
QY 184 TCRIVRQOYKOLSATKSDDTVAIPYGSRRHRLVLKGPDLHLYLTKTLOQTGKENSLS 243  
Db 152 tcrivrgskshvpekreenevaplgsrsvritvkgpahlfiesktlqgskgehsfns 211  
QY 244 TGFLVDSNSVDFOKFDPKELRMAGPLTADFIKIRNSGSADSTVOFIFQPIIHRWRE 303  
Db 212 pgvfventtvefgerqckfipgmadaffiktrytaakdsavvffqypishqwrq 271  
QY 304 TDFPSCATCGGGYQLTSAECYDLRSNRVADQYCHYYPENIKPKPLQECNLDPCPASD 363  
Db 272 tdfpctvtcggyqlsaecvdlrkrvvpdyhyyenvkpkpkkecsmdpcpsd 331  
QY 364 GYKOIMPYDLYHPLPRWEATPWTACSSCGGQTOSRAVSCVEEDIOGHVTSVEWKCMT 423  
Db 332 gfkempdyhfpplprwehnpwtacsvscggqldrrsfvcveesmhgeillqveewkmya 391  
QY 424 PKMPIAOPCNIFDCPKWLADEWSPCTVTCGGLRYRVVLCTIDHRGMITGCGSPKTKPHIK 483  
Db 392 pkpkvmtcnlfcdcpkwanewscvtcgrlryrvvclnhrghenvgonpqlkhhk 451  
QY 484 EECVPTPCYKPKKLEKLPVEAKLPKQAELEGAAYSEEFPIPEAWSACTVTCGVGTQ 543  
Db 452 eecvlpicykpkkspeaklpwkqaeleetriateptfipepsacsttcgpgvq 511  
QY 544 VRIVRCOVILFSQSADVLDFDECEGPKPASQ-----ACYAGPCSGEIPFENFDETDGL 598  
Db 512 vrevkcvllftqtelpeeececgkplpterpcilleacdespasreldiplp----- 565  
QY 599 FGLQDPDELYDEYEGFTKCSSECGGVQEAUVVSCNLKOTREPAENLGVTSRRPQLL 658  
Db 566 -----edsettydweyagftpctatclghqeatavclhctqctvnsldcmvhrpams 621  
QY 659 KSNLDPCPARWIGKWSPLCTGVLQTRDVFCSHLLSRENNETVILADELCRQPKPS 718  
Db 622 qacntepcprrwhvsgwpcsatcgvgqlrdvyclh-----pgetpappee-crdekph 675  
QY 719 TVOACNRFNCPWAYPAQWPCSTCGGQKREVLCQRMADGSLFELPFCPSASKPA 778  
Db 676 alqacqfcdpwpwhleewqgcrtcggtqnrvtvrtqlltdgfslnsdelcggpkas 735  
QY 779 COQACKKDCPSEWLLDTECSTSCGEGTQTRSAICRKMKTGLSTVWNSLCPPLPFS 838  
Db 736 shksartcdcpplhavgdwksvscvgvqgrkvcqqlaakgrriplesmmcdrlpgf 795  
QY 839 SSTRPCMLATCAR-PGRPSTK---HSPHIAAARKVYIQTTRQRLKHFVVGFAFYLKPKA 894  
Db 796 plvrscqmpcskkskcmktklgeqgpqllsvrvylqtreekrinitigsrayllpnts 855  
QY 895 VVLRCPARRVKPLITWEDKQHLISSTHTVAPFGVLKTHRLKPSDAGVYTSAGPARE 954  
Db 856 viikcpvrrfqlkwelwkgdrcqlqskrlgltksgslkhglaapdigvryciagsaqe 915  
QY 955 HFVYKILGGRKRLVARPLSPRSEEVLAHKGKGPKEALQTHKHQNGIFNSGSKAEKRGLA 1014  
Db 916 tvvlkligtndrliarp-----alrepmreygmhdhseanslgtvwhk--mrqmw 963  
QY 1015 ANFGSRY--DDLVSRL-----LLEQGGWPGELLASWFAQDSAEENNTTSEEDPGAEOVLH 1066  
Db 964 nnkndlylddhlsnqpfllrallghcnsagstnswelknkqfeavkvq---ga----- 1014  
QY 1067 LPFTWTEQRRLDDILGNLSQOPE--ELRDLYSKHLVAOLAQLFRSHLEHQDTLLKPSE 1124  
Db 1015 --ysmdtaq--fdeirnmqlmetgevsddlasqliyqivaelakapqthmq--wrgiq 1068  
QY 1125 RRTSPVTLSPHKHVSFGSSSLRTSSTGAGGGRRRPHRK-----PTILRKISAA 1173  
Db 1069 eetpp-----aaqlr-----getsvsgqshaknsqgkltfkpkqpvimrq-sqp 1111

QY 1174 QQLGASEVV--THLGOTVALASCTLSVLLHCRAIGHPRPTISWARGEEWQFSRLLQPD 1232  
Db 1112 psifnktinsrignvtvitrtevinilcdlitpseatytwkdgtllqpsvkiildgt 1171  
QY 1233 DSLQILAPVEADVGYTCNATNALGYDSVSIATVIAGKPLVKTSMTVINTEKPAVTVDI 1292  
Db 1172 gkigiqnptrkeggiyecsvanhlgdsvessvlyaeapvilsvernitkpehuhlsvv 1231  
QY 1293 GSTTKTVQGVVWTTNCQVAGVPEAEVWFRNKSIL-GSPHILHEGSLLLTNVSSSDQGLY 1351  
Db 1232 ggiyeaalganvtircpvgypqnitwkrsgslsgnvsllfngsllllgnvslenegy 1291  
QY 1352 SCRAANLHGETESTQLLIDPPQVPTOLEDIRA-----LLAATGPNLPSVLTSPLGTQLV 1407  
Db 1292 vciatnalgkavatsvhlhlerwpesrivflgghkkyilqatntrtns----- 1340  
QY 1408 LDPNSALLGCPKIKGHPVPNITWFGGQPIVTATGLTHHILAAAGQILQVANLSGSGQGEF 1467  
Db 1341 ndp-----tgeppp----- 1349  
QY 1468 SCLAQNEAGVLMQKASLVIOQYWNVSVDRLATCSASCGNRGVQOQRLRCLL-NSTEVPNAH 1526  
Db 1350 -----gepfwepgnwshcsatcghlgarigrpqcvmangqevseal 1390  
QY 1527 CAGKVRPAVQPIACNRDPCSRMWTWSACTRSCGGGVQTRRVTCOKLKASGISTPVSN 1586  
Db 1391 cdhlaqlagfepcniidrcparwftsvwscvsgcyshrgvtckrtkangtvqvvspp 1450  
QY 1587 DMCTQVAKRPVDTQACNQOQLCWEAFSSWGQCNGPCIGPHLAVQHRQVFCQTRDGTILPS 1646  
Db 1451 racap-kdrpigrkpcfgbpcvq--epgnrcpgrcmgravrmmqgrhtacqhns--d 1503  
QY 1647 EQCSALRPVSTQNCWEACSVHHRVSLWTLCATCGNYGFSQSRVCEVHARTKANPEH 1706  
Db 1504 snodrkrtlrrnctsgacdvcvhtgpkpctaagr-gfgrskvdcihtrscpkvavr 1562  
QY 1707 LCSGPRPANWQRNINTPCENNECRDTRYCEYKVKQLKCOLSOFKSRCCGTC 1759  
Db 1563 hcvqkkipiswrhc-lapscdrdetdtthymfvkhlnlesldrykqrccqsc 1614  
  
RESULT 7  
AAV87347  
ID AAV87347 standard; Protein; 643 AA.  
XX  
AC AAV87347;  
XX  
DT 11-MAY-2000 (first entry)  
XX  
XX Human signal peptide containing protein HSP-124 SEQ ID NO:124.  
KW Human; signal peptide-containing protein; HSP; diagnosis; cancer;  
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;  
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;  
KW reproductive disorder; developmental disorder; arteriosclerosis;  
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
KW Parkinson's disease; Huntington's disease; ovulatory defect;  
KW muscular dystrophy.  
XX  
OS Homo sapiens.  
XX  
XX W0200000610-A2.  
PN  
XX  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-US14484.  
XX  
PR 26-JUN-1998; 98US-0090762.  
PR 31-JUL-1998; 98US-0094983.  
PR 01-OCT-1998; 98US-0102686.

PR 11-DEC-1998; 98US-0112129.  
XX (INCY-) INCYTE PHARM INC.  
XX  
PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;  
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;  
PI Bandman O;  
XX  
XX WPI; 2000-160673/14.  
DR N-PSDB; AA298232.  
XX  
XX New human signal peptide-containing proteins useful in treatment,  
PT prevention and diagnosis of e.g. cancer, inflammation and  
PT cardiovascular disease  
XX  
XX Claim 1; Page 243-244; 327pp; English.  
XX  
XX AAZ98109 to AAZ98242 encode AA87224 to AA87357 which represent the  
CC human signal peptide-containing proteins HSPB-1 to HSPB-134. HSPBs have  
CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,  
CC neuroprotective, cardiovascular and antiasthmatic activities, and can  
CC be used in gene therapy. HSPBs can be used to treat or prevent disorders  
CC associated with decreased activity or function of HSPB. Antagonists of  
CC HSPB are used to treat or prevent disorders associated with increased  
CC activity or function of HSPB. Such diseases include cell proliferation  
CC (including cancer), inflammation, cardiovascular, neurological,  
CC reproductive or developmental disorders, (e.g. arteriosclerosis,  
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,  
CC asthma, Crohn's disease, microbial or other infections, congestive or  
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPB  
CC nucleic acids can be used for the recombinant production of HSPB, for  
CC detecting HSPB in standard hybridisation and amplification assays (for  
CC diagnosis and monitoring), in gene therapy, as antisense,  
CC triplex-forming or ribozyme therapeutics, for detecting related sequences  
CC or genetic variations, and for chromosomal mapping. HSPB are also used to  
CC raise specific antibodies (Ab) and to screen for agonists and  
CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
CC monitor, HSPB-related diseases (in usual immunoassays), as therapeutic  
CC antagonists, in competitive drug screens, and for purification of HSPB  
CC from natural sources.  
XX  
XX Sequence 643 AA;  
SQ

Query Match 37.2%; Score 3565; DB 21; Length 643;  
Best Local Similarity 100.0%; Pred. No. 3e-232;  
Matches 642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MECRRATPGTLLFLAFLLSSRTARSEEDRGLDAPGWPWSCRTCGGASYSLRRC 60  
Db 1 mecrratpgtllflafllssrtarseedrdglwdawgpwsecrctcggasyslrcc 60  
QY 61 LSSKSCGRNRYRTCSNVPDPPAGDFRAQCSAHNDVHHGQFYEWLPVSNPDNDPCS 120  
Db 61 lsskscegrnryrtcsnvpdppeagdfraqcsahndvhhghfyewlpvsnpdndpcs 120  
QY 121 LKQAKGTTLLWELAPKVLDTGRCYTESLDMCISGLCQIQVCGDHLGSTVKEDNCYCNCG 180  
Db 121 lkqakgtllwvelapkvldgtcrctesldmcisglcqiqvcdhqlgstvkedncycng 180  
QY 181 DGSFCLRVGRQYSQISLATSDDTVVAIPVGSRRHRLVLKPGDHLYLETKTLQGTGKENS 240  
Db 181 dgstclrvgrqysqlsatsddtvvaipvgsrrhrlvlkpgdhlyletktlqgtkgens 240  
QY 241 LSSGTGTLVNSVDFQKFPDKELRMAGPLTADFIKIRNSGADSTVQRFYQPIIHR 300  
Db 241 lssgtgtlvnsvdfqkfpdkelrmagpltdafikirnsagsadstvqrfyqpiih 300  
QY 301 WRETDFFPCSATCGGGYQLTSAECYDLRSNRVADQYCHYYPENIKPKKLQECNLDPCP 360  
Db 301 wretdffpcsatcggyqltsaecydlrsnrvadqychyyphenikpkklqecnldpcp 360

QY 361 ASDGYKOIMPYDLYHPLRWEATPWTACSSSCGGGIOSRAVSCVVEEDIQGHVTSVEWKC 420  
Db 361 asdgykqimpydlyhplrweatpwtacssscgggiqsraavscveediqghvtsveewkc 420  
QY 421 MYTPKMPDIAOPCNIFDCPKWLAQEWSPCTVTCGGLRVVVLCTIDHRGHTGGSPKTKP 480  
Db 421 mytpkmpdiapcniifdcpkwlagewspctvtcggglryvvlcidhrgmhtggspktkp 480  
QY 481 HIKEECIVPTPCYKPKKLPVEAKLPWFKQAELEEGAAVSEEPSFIPEAWSACTVTTCGV 540  
Db 481 hikeecivptpcykpkklpveaklpwfkaqlpveakleegaavseepsfipeawsactvtcgv 540  
QY 541 GTQVRIVRQCVLLSFSQSVADLPIDCEGPKPASQACYAGPCSGEIPENPDFTDGLFG 600  
Db 541 gtqvrivrqcvllsfsqsvadlpidecegpkpasqacypasgracyagpcsgelpefnpdetdglfg 600  
QY 601 GLQDFDELDFWEYEGFTKCSFSCGGGVQEAUVVSLNKOTREP 642  
Db 601 glqdfdeldfweyegftkcsescgggvqeaavvslcnkqtrep 642  
RESULT 8  
ABBI0454  
ID ABB10454 standard; Protein; 645 AA.  
XX ABB10454;  
AC ABB10454;  
DT 10-JAN-2002 (first entry)  
XX Human cDNA SEQ ID NO: 762.  
DE Human; gene therapy; neural disorder; immune system disorder;  
XX muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
KW proliferative disorder; inflammation.  
XX Homo sapiens.  
OS  
XX W0200154474-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01349.  
XX  
XX 31-JAN-2000; 2000US-179065P.  
PR 04-FEB-2000; 2000US-180628P.  
PR 24-FEB-2000; 2000US-184664P.  
PR 02-MAR-2000; 2000US-186350P.  
PR 16-MAR-2000; 2000US-189874P.  
PR 17-MAR-2000; 2000US-190076P.  
PR 18-APR-2000; 2000US-198123P.  
PR 19-MAY-2000; 2000US-205515P.  
PR 07-JUN-2000; 2000US-209467P.  
PR 28-JUN-2000; 2000US-214886P.  
PR 30-JUN-2000; 2000US-215135P.  
PR 07-JUL-2000; 2000US-216647P.  
PR 07-JUL-2000; 2000US-216880P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217496P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.  
PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225213P.  
PR 14-AUG-2000; 2000US-225214P.  
PR 14-AUG-2000; 2000US-225266P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 14-AUG-2000; 2000US-225758P.

```
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 08-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231444P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.

PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 06-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPT; 2001-476161/51.
XX N-PSDB; ABA06676.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
XX polypeptide is used in preventing, treating or ameliorating a medical
XX condition -
XX
XX Claim 11; SEQ ID NO: 762; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
XX DNAs. These can be used in the treatment of neural, immune system,
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
XX renal and proliferative disorders and inflammation. The present sequence
XX is a protein of the invention.
XX
XX Sequence 645 AA;
SQ
Query Match 35.7%; Score 3423; DB 22; Length 645;
Best Local Similarity 99.8%; Pred. No. 1.2e-222;
Matches 636; Conservative 1; Mismatches 0; Indels 0; Caps 0;
QY 1126 RTSPVTLSPHKHVSFGFSSSLRTSTGDAGGSRPRPKPTILRKISAAQOLSASEVVTHL 1185
Db 9 rtsptlspkhkvsfgfssslrtsstgdagggrrprpkpkillrkisaaqqlsasevvthl 68
QY 1186 GQTVLASGTLVLLHCEAIGHPRPTISWARGNEVQFSDRILLQPDSDLIQLAPVEADV 1245
Db 69 gqtvalasgtlsvllhceai ghprptiswargneevqfsdrillqpdssliqlapveadv 128
QY 1246 GFYTCNATNALGYDSVSIATVTLACKPLVKTSRMTVINTEKPAVTVDIGSTIKVQGVNVT 1305
Db 129 gfytcnatanlgydsvsiavtvtlackplvktksrmtvintekpavtvdigstiktvggvnt 188
```

```
QY 1306 INCOVAGPEAEVTFWRPNKSKIGSPHLLHSGSLLTNNVSSDOGLYSCTAANLHGELTES 1365
Db 189 incqavagpeaevtfwrpnkskigspshllhsgslttcnvssddgylsctaanhgелtes 248
QY 1366 TQLLIIDPQVPTQLEEDIRALLAATGPNLPVLTSLTSLGTQLVLDPGNSALLGCPKIGHPV 1425
Db 249 tqllilddpqpvtqleedirallaatgpnlpvltspigtqlvldpgnsallgcpikghpv 308
QY 1426 PNITWFHGGQPIYVATGLTHHILAAQIILQVANLGGSGQEFSCLAQNEAGVLMQKASLV 1485
Db 309 pnitwfhgqgpiviatglthhilaagqilqvanlsgsgqefscslaqaagvlmqkaslv 368
QY 1486 IQDYVWSVDRLATPCASCGNRGVQOQRLRLCLLNSTEVNPAHCAGKVPVAVOPIACNRDDC 1545
Db 369 iqdyvwsvdrlatpcasacnrgvqprrlcllnstevnpahcagkvrpavqpiacnrddc 428
QY 1546 PSRWMTWSACTRSCGGGVQTRVTCQKLIKASGISITPPVNDMCTQVAKRPVDTQACNQ 1605
Db 429 psrwmtwsactrscgggvqtrvtcqklikasgistpvsndmctqvakrpvdtqacnq 489
QY 1606 LCVEWAFSSWGQCNIGCPHGLAVQHRQVFCQTRDGIITLPEQCSCALPRPVSTQNCWSEA 1665
Db 489 lcvewafsswgqcnigcpvghlavqhrqvfctrdgiltlpeqscsalprpvstqncwsea 548
QY 1666 CSVHWRVSLWTLCTATCGNYGFSRRVECVHARTNKAVPHELCSWGPANWQRNITPC 1725
Db 549 cshwrvslwtlctatcgnvgfrrvecvhartnkavpehlcswgprpanwqrnctipc 608
QY 1726 ENMECDRTTRYCKVKQKLCQLSQPKSRCCGCGKA 1762
Db 609 enmecdrttryckvkqkicqisqfksrccgctgka 645

RESULT 9
AA41729
ID AAY41729 standard; Protein; 525 AA.
XX
AC AAY41729;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human P101071 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079234.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH ) GENENTECH INC.
XX
```

PI	Wood WL, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;	DT	18-DEC-2001 (first entry)	XX
DR	WPI; 1999-551358/46.	XX	Human PRO polypeptide sequence #218.	XX
DR	N-PSDB; AAZ34146.	XX	PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;	XX
XX		KW	dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;	KW
PI	New secreted and transmembrane polypeptides and their polynucleotides,	KW	blood; chondrocyte cell; cell proliferation; cell differentiation; colon;	KW
PT	useful for treating blood coagulation disorders, cancers and cellular	XX	adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.	XX
PT	adhesion disorders -	OS	Homo sapiens.	OS
XX	Claim 12; Fig 120; 530pp; English.	XX	WO200168848-A2.	XX
XX	The present invention describes secreted and transmembrane polypeptides	PN	20-SEP-2001.	XX
CC	and their polynucleotides. The nucleotide sequences are useful as	XX	28-FEB-2001; 2001WO-US06520.	XX
CC	sources of probes, primers, for chromosome mapping, and for generation	XX	01-MAR-2000; 2000WO-US05601.	XX
CC	of antisense sequences. They can also be used to create transgenic	PR	02-MAR-2000; 2000WO-US05841.	PR
CC	animals. The proteins can be used to treat a variety of diseases and	PR	03-MAR-2000; 2000US-187202P.	PR
CC	disorders, depending on their function. Diseases that may be treated	PR	06-MAR-2000; 2000US-186968P.	PR
CC	include blood coagulation disorders, cancers and cellular adhesion	PR	14-MAR-2000; 2000US-189320P.	PR
CC	disorders. They may also be used to raise antibodies. AAZ33891 to	PR	14-MAR-2000; 2000US-189328P.	PR
CC	AA234338, and AAY41685 to AAY41774 represent polynucleotide and	PR	15-MAR-2000; 2000WO-US06884.	PR
CC	polypeptide sequence given in the exemplification of the present	PR	21-MAR-2000; 2000US-190828P.	PR
CC	invention.	PR	21-MAR-2000; 2000US-191007P.	PR
XX	Sequence 525 AA:	PR	21-MAR-2000; 2000US-191048P.	PR
SQ		PR	21-MAR-2000; 2000US-191314P.	PR
		PR	28-MAR-2000; 2000US-192655P.	PR
		PR	29-MAR-2000; 2000US-193032P.	PR
		PR	29-MAR-2000; 2000US-193053P.	PR
		PR	30-MAR-2000; 2000WO-US08439.	PR
		PR	04-APR-2000; 2000US-194449P.	PR
		PR	04-APR-2000; 2000US-194647P.	PR
		PR	11-APR-2000; 2000US-195975P.	PR
		PR	11-APR-2000; 2000US-196000P.	PR
		PR	11-APR-2000; 2000US-196187P.	PR
		PR	11-APR-2000; 2000US-196690P.	PR
		PR	11-APR-2000; 2000US-196820P.	PR
		PR	18-APR-2000; 2000US-198121P.	PR
		PR	18-APR-2000; 2000US-198585P.	PR
		PR	25-APR-2000; 2000US-199397P.	PR
		PR	25-APR-2000; 2000US-199550P.	PR
		PR	25-APR-2000; 2000US-199654P.	PR
		PR	03-MAY-2000; 2000US-201516P.	PR
		PR	17-MAY-2000; 2000WO-US13705.	PR
		PR	22-MAY-2000; 2000WO-US14042.	PR
		PR	30-MAY-2000; 2000WO-US14941.	PR
		PR	02-JUN-2000; 2000WO-US15264.	PR
		PR	05-JUN-2000; 2000US-209832P.	PR
		PR	28-JUL-2000; 2000WO-US20710.	PR
		PR	22-AUG-2000; 2000US-0644848.	PR
		PR	24-AUG-2000; 2000WO-US23328.	PR
		PR	08-NOV-2000; 2000WO-US30952.	PR
		PR	01-DEC-2000; 2000WO-US32678.	PR
		PR	20-DEC-2000; 2000WO-US34956.	PR
		XX	(GETH ) GENENTECH INC.	XX
		PA	Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;	XX
		PI	Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;	PI
		XX	WPI; 2001-602746/68.	XX
		XX	N-PSDB; AAS46142.	XX
		XX	Novel nucleic acids encoding PRO polypeptides, used to diagnose the	XX
		PT	presence of tumours, such as prostate and breast tumours, in mammals and	PT
		PT	to screen for modulators of the compounds -	PT
		XX	Claim 11; Fig 436; 774pp; English.	XX
		PS	Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.	PS
		XX	The PRO polypeptides and their associated nucleic acids can be used to	XX
		CC		CC

CC detect the presence of a tumour in a mammal by comparing the level of  
CC expression of a PRO polypeptide in a test sample of cells from the animal  
CC and a control sample of normal cells, whereby a higher level of  
CC expression in the test sample indicates the presence of a tumour in the  
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
CC and rabbits but are preferably human. The polypeptides can be used to  
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,  
CC when contacted with it. A specific polypeptide can be used to stimulate  
CC the proliferation or differentiation of chondrocyte cells. The PRO  
CC proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
CC can be used for genetic analysis of individuals with genetic disorders.

XX Sequence 525 AA;

Query Match 30.4%; Score 2917; DB 22; Length 525;  
Best Local Similarity 99.8%; Pred. No. 1.4e-188;  
Matches 524; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60  
DB 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60  
QY 61 LSSKSCGGRNIRYTCNVDCPPEAGDFRAQOCSAHNDVKHHGOFYEWLPVSNDDPNPCS 120  
DB 61 LSSKSCGGRNIRYTCNVDCPPEAGDFRAQOCSAHNDVKHHGOFYEWLPVSNDDPNPCS 120  
QY 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180  
DB 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180  
QY 181 DSGTCLVRCQYKSQLSATKSDTVAIPYGSRHRLVLKGPDLHYLETKTQCTKGENS 240  
DB 181 DSGTCLVRCQYKSQLSATKSDTVAIPYGSRHRLVLKGPDLHYLETKTQCTKGENS 240  
QY 241 LSGTGFVLVDSNVDFQKPPDKELRMAGPLTADFIVKIRNSGASDSTVOFIYQPTIHR 300  
DB 241 LSGTGFVLVDSNVDFQKPPDKELRMAGPLTADFIVKIRNSGASDSTVOFIYQPTIHR 300  
QY 301 WRETDFPSCATCGGGYQLTSECYDLNRNVVADYCHYPENIKPKKLECNLDPCP 360  
DB 301 WRETDFPSCATCGGGYQLTSECYDLNRNVVADYCHYPENIKPKKLECNLDPCP 360  
QY 361 ASDGYKQIMPYDLYHPLPWEATPWTACSSCGGGIQRSAVSCVEEDIOGHVTSVEEWC 420  
DB 361 ASDGYKQIMPYDLYHPLPWEATPWTACSSCGGGIQRSAVSCVEEDIOGHVTSVEEWC 420  
QY 421 MYTPKMPIAQCNIIFDCPKWLAQWSPCTVTCGGGLRYRWVLCIDHRGHTGSCSKTKP 480  
DB 421 MYTPKMPIAQCNIIFDCPKWLAQWSPCTVTCGGGLRYRWVLCIDHRGHTGSCSKTKP 480  
QY 481 HKEECIVTPTCYKPKKLPVAKLIPWFKQAQLEEGAAVSEEPS 525  
DB 481 HKEECIVTPTCYKPKKLPVAKLIPWFKQAQLEEGAAVSEEPS 525

RESULT 11  
AAB44285  
ID AAB44285 standard; Protein; 525 AA.

XX AAB44285;  
XX AAB44285;  
DT 08-FEB-2001 (first entry)  
XX Human PRO1071 (UNQ528) protein sequence SEQ ID NO:301.  
XX Human; secreted protein; transmembrane protein; PRO; EST; cytotstatic;  
KW expressed sequence tag; detection; cancer.  
XX Homo sapiens.

XX WO200053756-A2.  
XX 14-SEP-2000.  
XX 18-FEB-2000; 2000WO-US04341.  
XX 08-MAR-1999; 99WO-US05028.  
XX 12-MAR-1999; 99US-0123957.  
XX 29-MAR-1999; 99US-0126773.  
XX 21-APR-1999; 99US-0130232.  
XX 28-APR-1999; 99US-0131445.  
XX 14-MAY-1999; 99US-0134287.  
XX 23-JUN-1999; 99US-0141037.  
XX 26-JUL-1999; 99US-0145698.  
XX 29-OCT-1999; 99US-0162506.  
XX 30-NOV-1999; 99WO-US28313.  
XX 02-DEC-1999; 99WO-US28551.  
XX 02-DEC-1999; 99WO-US28565.  
XX 16-DEC-1999; 99WO-US30095.  
XX 30-DEC-1999; 99WO-US31243.  
XX 30-DEC-1999; 99WO-US31274.  
XX 05-JAN-2000; 2000WO-US00219.  
XX 06-JAN-2000; 2000WO-US00277.  
XX 06-JAN-2000; 2000WO-US00376.  
XX (GETH ) GENENTECH INC.  
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
XX Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
XX Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
XX Klijavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
XX Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;  
XX WPI: 2000-611443/58.  
XX N-PSDB; AAC78532.

Novel PRO polypeptides and polynucleotides used in detection methods,  
PT to target bioactive molecules to specific cells, and to modulate  
PT cellular activities -  
XX Claim 12; Fig 120; 636pp; English.

XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
CC sequence tag) sequences which encode secreted or transmembrane PRO  
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
CC activity. The polynucleotides and polypeptides can be used for detecting  
CC the presence of PRO polypeptides in samples, for linking bioactive  
CC molecules to cells and for modulating biological activities of cells,  
CC using the polypeptides for specific targeting. The polypeptide targeting  
CC can be used to kill the target cells, e.g. for the treatment of cancers.  
CC The polypeptide pairs provide specific targeting of bioactive molecules  
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in  
CC the isolation of the PRO polynucleotide sequences.

XX Sequence 525 AA;

Query Match 30.3%; Score 2907; DB 21; Length 525;  
Best Local Similarity 99.4%; Pred. No. 6.8e-188;  
Matches 522; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60  
DB 1 MECCRRATPGTLLFLAFLLSSRTARSEDRDGLWDAGPWSFCSRTCCGGASYSILRRC 60  
QY 61 LSSKSCGGRNIRYTCNVDCPPEAGDFRAQOCSAHNDVKHHGOFYEWLPVSNDDPNPCS 120  
DB 61 LSSKSCGGRNIRYTCNVDCPPEAGDFRAQOCSAHNDVKHHGOFYEWLPVSNDDPNPCS 120  
QY 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180  
DB 121 LKQAKGTTLVVELAPKVLDTGTCRYTESLDMCISGLCQIVGCDHQLGSTVKEDNCVCNG 180



QY 181 DGSTCLVRGQYKQSLSATKSDTVAIPYSGRHRIRLVKGPDLHYLETKTLOGTKGENS 240  
Db 181 dgstclvrqyksqslsatksdttvvalpygsrhrirvlkghpdhlyletktlqgkgens 240  
QY 241 LSSTGTLVDNSSVDFQFPDKEILRMAGPLTADFTIVKIRNSGSADSTVQFIYQPIIHR 300  
Db 241 lsstgtflvndssvdfqfpdkeilrmagpltdftivkirnsgsadstvqfifyqpiih 300  
QY 301 WRETDFPFCATCGGQYQLTSAECYDLRNRVVADQYCHYYPENIKPKLOECNLDPCP 360  
Db 301 wretdfpfcatscgggyqltsaecydlrnrvvadqychyypenlkpkloecnldpcp 360  
QY 361 ASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVSCVEEDIQGHVTSVEEWKC 420  
Db 361 asdgykqimpydlyhplprweatpwtacsscggiqsravscveedighvtsveewkc 420  
QY 421 MYTPKMPIAQCNIIFDCPKWLAQESPCVTCTCGGLRYRVVLCIDHRGMHTGCCSPKTKP 480  
Db 421 mytpkmpiaqcniifdcpkwlaqespcvtctcgglryrvvlicidhrgmhtgcsptktp 480  
QY 481 HIKEECIVPTPCYKPKELPVEAKLPWFKQAOELEGAASVEEPS 525  
Db 481 hkeeciivptpcykpkelpveaklpwfkaqealeegaavseeps 525  
RESULT 12  
AAB72290  
ID AAB72290 standard; Protein; 525 AA.  
AC AAB72290;  
XX  
DT 14-MAY-2001 (first entry)  
DE Human ADAMTS-R1 amino acid sequence.  
XX  
KW ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;  
KW tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;  
KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;  
KW metastasis; embryogenesis; egg implantation; ADAMTS-R1.  
XX  
OS Homo sapiens.  
XX  
PN WO200111074-A2.  
XX  
PD 15-FEB-2001.  
XX  
PF 03-AUG-2000; 2000WO-US21223.  
XX  
PR 06-AUG-1999; 99US-0369364.  
XX  
PA (CLEV-) CLEVELAND CLINIC FOUND.  
PA (APTE/) APTE S S.  
PA (HURS/) HURSKAINEN T L.  
PA (HIRO/) HIROHATA S.  
XX  
EI Apte SS, Hurskainen TL, Hirohata S;  
XX  
DR WPI; 2001-159978/16.  
DR N-PSDB; AAF63447.  
XX  
PT Murine and human 'A Disintegrin-like And Metalloprotease domain with  
PT Thrombospondin type I motifs' proteins and the nucleic acids encoding  
PT them, useful for treating e.g. tumours, inflammation and arthritis -  
XX  
PS Claim 1; Fig 11; 181pp; English.  
XX  
CC This invention relates to murine and human ADAMTS-N (A disintegrin-like  
CC and metalloprotease domain with thrombospondin type I motifs) proteins,  
CC designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the  
CC invention are cDNA sequences encoding the proteins, and antibodies  
CC specific for the proteins. The nucleic acid sequences and proteins may be  
CC used in the prevention, diagnosis and treatment of diseases associated

CC with inappropriate ADAMTS-N expression. Disorders that may be treated  
CC using the nucleic acids, proteins and antibodies include, for example  
CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos  
CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage  
CC in arthritic (both inflammatory and non-inflammatory) disease,  
CC angiogenesis, tumour growth and metastases, and they may also be used for  
CC controlling embryogenesis and implantation of fertilised eggs. The  
XX present sequence represents human ADAMTS-R1.  
SQ Sequence 525 AA;  
Query Match 30.3%; Score 2907; DB 22; Length 525;  
Best Local Similarity 99.6%; Pred. No. 6.8e-188;  
Matches 523; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MECCRRATPGTLLFLAFLLLSSRTARSEEDRDGLWDAMPWSECSTRTCGGASYSILRRC 60  
Db 1 meccrratpgtllflaflllsrtarseedrdglwdawgpwsecsttcgggaansilrrc 60  
QY 61 LSKSCCEGRNIYRTCSNVDCPEAGDFRAQCCSAHNDVKHHGQFYEWLPVSNDDPNPNCs 120  
Db 61 lsksccegrniyrctcsnvdcppeagdfraqccsahndvkhgqfyewlpvsnddpnpcs 120  
QY 121 LKQAKGTTLVVELAPKVLDTGRTCYTESLDMGICSLQCIVGCDHOLGSTVKEDNCQVCNG 180  
Db 121 lkqakgttlvvelapkvldgtrcritesldmcisglcqiavgcdhqlgstvkedncqvcng 180  
QY 181 DGSTCLVRGQYKQSLSATKSDTVAIPYSGRHRIRLVKGPDLHYLETKTLOGTKGENS 240  
Db 181 dgstclvrqyksqslsatksdttvvaipygsrhrirvlkghpdhlyletktlqgkgens 240  
QY 241 LSSTGTLVDNSSVDFQFPDKEILRMAGPLTADFTIVKIRNSGSADSTVQFIYQPIIHR 300  
Db 241 lsstgtflvndssvdfqfpdkeilrmagpltdftivkirnsgsadstvqfifyqpiih 300  
QY 301 WRETDFPFCATCGGQYQLTSAECYDLRNRVVADQYCHYYPENIKPKLOECNLDPCP 360  
Db 301 wretdfpfcatscgggyqltsaecydlrnrvvadqychyypenlkpkloecnldpcp 360  
QY 361 ASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVSCVEEDIQGHVTSVEEWKC 420  
Db 361 asdgykqimpydlyhplprweatpwtacsscggiqsravscveedighvtsveewkc 420  
QY 421 MYTPKMPIAQCNIIFDCPKWLAQESPCVTCTCGGLRYRVVLCIDHRGMHTGCCSPKTKP 480  
Db 421 mytpkmpiaqcniifdcpkwlaqespcvtctcgglryrvvlicidhrgmhtgcsptktp 480  
QY 481 HIKEECIVPTPCYKPKELPVEAKLPWFKQAOELEGAASVEEPS 525  
Db 481 hkeeciivptpcykpkelpveaklpwfkaqealeegaavseeps 525  
RESULT 13  
AAE07868  
ID AAE07868 standard; Protein; 845 AA.  
XX  
AC AAE07868;  
XX  
DT 01-NOV-2001 (first entry)  
DE Novel human protein (NHP) #6.  
XX  
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;  
KW cytoskeletal; antisense therapy; angiogenesis; biological disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200161011-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 15-FEB-2001; 2001WO-US05290.





Db 361 tnvteqrrlddlnlsgqpeeirdlyskhnlvaqlaqeifrshlehqdtllkpxerittp 420  
QY 1130 VTLSPKHVSGFSSSLRTSSSTGDAGGSRPHRKPITILRKISAAOOLSASEVVTHLGQTV 1189  
Db 421 vtlspkhvsgfssslrtsstgdaggsrrphrkptliirkisaagqlsasevvthlgqlv 480  
QY 1190 ALASGTLSV 1198  
Db 481 alasgtlsv 489

Search completed: July 24, 2002, 04:17:42  
Job time: 7030 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 22:39:31 : Search time 6233.88 seconds  
(without alignments)  
17754.665 Million cell updates/sec

Title: US-10-044-807-1  
Perfect score: 5289  
Sequence: 1 atggaatgctgcctcgggc.....gaacttggaagcgtga 5289

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.or.\*

21: em.ov.\*

22: em.pat.\*

23: em.ph.\*

24: em.pl.\*

25: em.pr.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

RESULT 1

AF176313

LOCUS

DEFINITION

AF176313

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

AF176313

1571.4

29.7

1803

9

AF176313

1571.4

29.7

1803

9

AF176313

1571.4

29.7

1803

9

AF176313

1571.4

29.7

1803

9

AF176313

1571.4

29.7

1803

9

ALIGNMENTS

AF176313 1803 bp mRNA linear PRI 08-AUG-2001  
Homo sapiens ADAM-TS related protein 1 (ADAMTSR1) mRNA, complete  
cds.

AF176313 GI:15099920

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1803)

Hirohata, S. and Apte, S.S.

A novel member of ADAM-TS related gene, ADAM-TS1 (A

disintegrin-like And Metalloproteinase domain with Thrombospondin

type I modules Related gene-1)

Unpublished

2 (bases 1 to 1803)

Hirohata, S. and Apte, S.S.

Direct Submission

Submitted (06-AUG-1999) Biomedical Engineering, The Cleveland

Clinic Foundation, 9500 Euclid Avenue, ND20, Cleveland, OH 44195,

Phandology N-term

```
FEATURES             US          Location/Qualifiers
source                1..1803
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
gene                 1..1803
                     /gene="ADAMTSR1"
                     /note="a disintegrin-like and metalloproteinase domain
                     with thrombospondin type I modules related gene-1"
CDS                  51..1628
                     /gene="ADAMTSR1"
                     /note="ADAM-TSRI; does not contain metalloproteinase
                     domain"
                     /codon_start=1
                     /product="ADAM-TS related protein 1"
                     /protein_id="AAK84170.1"
                     /db_xref="GI:15099921"
                     /translation="MECCRRATPGTLLFLAFLLSSRTARSEDRDLWDAMPWSE
                     CSRTGGGAANSLRCLSKSCSEGRNRYTCNSVDCPEAGDFRAQCSAHNDVXKH
                     GQFYEMLPVSNDFNDPCLSKCAQKQATLIVELAPKVDLDRGCTESLDMCSIGLQIV
                     GDHQHGSTVKEDNCVNGDSCTRLVGOYKSLQSAKSDSDVVAIPYSGSRHRLV
                     LKQPHLYLETLLQGTGKNSLSSTGTLVDNSSYDFQKFPDKELRMAGPSTADFI
                     VKIRSGSDSTVQFIYQPIIHRWRETDFPCSATCGGQYQLTSAECYDLRNRVVA
                     DOYCHYYPINIKPKLOCNLDPCPASDGYKOIMPYDLYHPLPRWEATPWATCSSC
                     GGQIOSRAVSCVEEDTQGHVTSVEWKMYTPKMPAQPCNLPDCPKWLAQWSPCTV
                     TCGQLRYRVVICIDHRGMHTGCCSPKTPHKEECIVPPCYKPKKULPVEAKLPWF
                     KQAQLEEGAAVSEEPS"
                     /variation
                     /gene="ADAMTSR1"
                     /note="changes translation from ala-asn to ser-tyr"
                     /replace="tcttao"
BASE COUNT          475 a 460 c 457 g 411 t
ORIGIN
1 atggaatgctgcgcgtcggaactctcggcacactgctctctcttcttgggtttctgctc 60
2 51 ATGGAATGCTGCCTCGGGCAACCTCTGGCACACTGCTCTCTTCTTGGGTTCTCTGCTC 110
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

```
Db 531 GGCTGGGATCACCAGCTGGGAAGCACCGTCAAGGAAGATAACTGTGGGGTCTGCAACGGA 590
QY 541 gatgggtccacctccggctggttcaggagggcagtaataatcccgctctccgcaacacaa 600
Db 591 GATGGGTCCACCTCGCGCTGGTCCGAGGGCAGTATAAATCCAGCTCTCCGCAACCAAAA 650
QY 601 tcgatgatactggttggaattccctatggaagtgaacatatcgccctgtgttataaaa 660
Db 651 TCGGATGATAGTGTGGTGTGCAATTCCTATGGAAGTAGACATATTCGCTTGTCTTAAAA 710
QY 661 ggtcctgatacttatactctggaacacaaacccctccaggaggaactaaagtggaacagt 720
Db 711 GGTCTGATCACTTATATCTGGAACCAAAACCCCTCCAGGGGACTAAAGGTGAACACAGT 770
QY 721 ctcagctccacaggaactttcttctgtggacaattctagtgtggactccagaatttcca 780
Db 771 CTCAGCTCCACAGCAACTTTCTTCTGTGGACAATCTAGTGTGGACTTCCAGAAATTTCCA 830
QY 781 gacaaagagatactgagaatgctggaccactcagcagagatttcaattgtcaagattcgt 840
Db 831 GACAAAAGAGATAC7GAGAATGGCTGGACCACTCAACAGAGATTTCATTGTCAAGATTGCT 890
QY 841 aactcgggctccgctgacagtcacagtcacagtcacagtcacagtcacagtcacagtcac 900
Db 891 AACTCGGGCTCCGCTGACAGTACAGTCCAGTTCATCTTCATCAACCCATCATCCACGGA 950
QY 901 tggaggagagcggattttcttctgtcagcaacacctgtgagaggttatcagctgaca 960
Db 951 TGGAGGGAGAGCGGATTTCTTCTTGTCTCAGCAACCTGTGGAGGAGGTATCAGCTGACA 1010
QY 961 tcgctgagtgctacgatactgagaggaacccgtggtgtgtgtgacccaatactgtcaat 1020
Db 1011 TCGGCTGAGTGCTACGATCTGAGGAGCAACCGTGTGGTGTGCTGACCAATACATGTCAT 1070
QY 1021 taaccagagaacatacaaacacacacacacacacacacacacacacacacacacacacac 1080
Db 1071 TACCCAGAGAACATCAAAACCCAAACCCAAAGCTTCAGGAGTCCAACTTGATTCCTTGTC 1130
QY 1081 gccagtacggatacaagcagatcatgctcttatgacctctaccatcccccttccctggtg 1140
Db 1131 GCCAGTGACGGATACAGCAGATCATGCTTATGACCTCTACCATCCCTTCCCTCGGTGG 1190
QY 1141 gagccaccccaatgagccgctgctctctctctctctctctctctctctctctctctctct 1200
Db 1191 GAGGCCACCCCATGGACCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1250
QY 1201 gtttctctgttggaggagacatccaggggcagtcacattcagtggaagagtggaatgc 1260
Db 1251 GTTTCTCTGTGTGGAGGAGACATCCAGGGGCAATGTCATTTCAGTGGAGAGTGGAAATGC 1310
QY 1261 atgtacacccctaaagatgcccctcgcagccctgcaacatttttgactgcccctaaatgg 1320
Db 1311 ATGTACACCCCTAAGATGCCCATCGCGAGCCCTGCAACATTTTGTACTGCCCTAAATGG 1370
QY 1321 ctgacacagagtggtctccgtgacagtgacatgagcagtgagcagcctcagatccgtgtg 1380
Db 1371 CTGCAACAGGAGTGTCTCCGTGACAGTGACGTGTGCGCAGGGCCCTCAGATACCGTGTG 1430
QY 1381 gtcctctgcatcgaccatcgagggaatgcacacagagagctgtagcccaaaaaaacagccc 1440
Db 1431 GTCCTCTGTCATCGACCATCGAGGAATGCACACAGAGAGGCTGTAGCCCCAAAAACAAAGGCC 1490
QY 1441 cacataaaagaggaatgcatactgatacccaactccccctgctataaaccacaaagagaactcca 1500
Db 1491 CACATAAAAGAGGAATGCATCGTACCCACTCCCTGCTATAAACCCAAAGAGAAACTTCCA 1550
QY 1501 gtcagagccagtggtgcaatggttcaacaagctcaagagctagagagagagctgctgtg 1560
Db 1551 GTCAGGCGCAGTTGCCATGTTTCAACAAAGCTCAAGAGCTAGAGAGAGAGAGCTGCTGTG 1610
QY 1561 tcagagagagccctgcttcaatccacagagcgtggtcgg 1597
Db 1611 TCAGAGGAGCCCTCGTAAGTTGTAAAGACACAGACTG 1647
```







LOCUS AC011701 205510 bp DNA linear PRI 27-FEB-2001  
 DEFINITION Homo sapiens 12 BAC RP11-937121 (Roswell Park Cancer Institute  
 Human BAC Library) complete sequence.  
 AC011701  
 AC011701.22 GI:13129402  
 HGK.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 205510)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alzbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,  
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bontie,S., Brileva,M., Brown,B., Brown,M., Bryant,N.P., Buhay,C.,  
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Gavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chiu,D., Chowdhry,I., Christophoulos,C., Cleveland,C.D.,  
 Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Emerling,S., Escotio,M., Falls,T., Ferraguto,D.,  
 Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,  
 Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,  
 Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,  
 Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,  
 Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsí,F.,  
 Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Lewisq,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,  
 Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,  
 Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,  
 Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,  
 Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,  
 Ogih,M., Okwuonu,G., Oraguene,N., Oviedo,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,M., Ren,G., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,  
 Shoostari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,  
 Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,  
 Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,  
 Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,  
 Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,  
 Washington,C., Watlington,S., Williams,G., Williamson,A.,  
 Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,  
 Zorrilla,S., Kucherlapati,R. and Gibbs,R.  
 Direct Submission  
 2 (bases 1 to 205510)  
 Unpublished  
 2 (bases 1 to 205510)  
 Worley,K.C.  
 Direct Submission  
 Submitted (12-OCT-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 205510)  
 Worley,K.C.  
 Direct Submission  
 Submitted (27-FEB-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Feb 27, 2001 this sequence version replaced gi:13096014.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT-----

#### ----- Summary Statistics -----

Contig length: 205510  
 Phrap values in estimate: 203901  
 Average error rate (BCM-phrap estimate): 0.000164184  
 Fraction of Phrap values less than 40 : 0.0264246  
 Number of consensus changing edits: 29  
 Number of N's in consensus : 0

#### ----- Consensus changing edits -----

Position	Original*Context	Edited*Context
14478	taaaagccag(n)aaactatat	taaaagccag(g)aaactatat
14480	aaagccagga(a)aaactatagt	aaagccagga(a)aaactatagt
18544	tccttgaga(n)nnctgccagt	tccttgaga(g)agctgcagt
18545	ccttgagag(n)nnctgccagt	ccttgagag(a)gctgcagt
18546	cttgagann(n)ctgccagtgg	cttgagaga(g)ctgcagtgg
30655	gccagagca(n)cagtcacaaat	gccagagca(a)cagtcacaaat
48971	taaaacactt(n)tgctgcccc	taaaacactt(c)tgctgcccc
85905	ctgggttg(g)ttttttttt	ctgggttg(t)ttttttttt
87055	aaacatgaat(c)aaacataca	aaacatgaat(c)aaacataca
96536	ccattttc(n)ttttttttt	ccattttc(t)ttttttttt
124767	ggcaaataca(n)nnactctgt	ggcaaataca(c)nnactctgt
124768	gcaaatacan(n)nnactctgtt	gcaaatacan(c)nnactctgtt
124769	caaatacann(n)acctctgtc	caaatacann(c)acctctgtc
124849	gccacacac(n)caaatgtaca	gccacacac(c)caaatgtaca
125377	ttatagtcta(g)aaaggaaga	ttatagtcta(g)aaaggaaga
125378	aattctcttg(g)gaggaacaa	aattctcttg(t)gaggaacaa
128124	ccagagaggt(n)ttgttcaga	ccagagaggt(c)ttgttcaga
128126	agagaggtnt(n)ttttacgac	agagaggtt(c)ttttacgac
128206	tgagggtaca(n)agctagtgg	tgagggtaca(g)agctagtgg
136767	gctcaagitt(n)caggtagaaa	gctcaagitt(c)caggtagaaa
153400	tcagctcttc(n)tcactctagc	tcagctcttc(c)tcactctagc
159220	cccaaggcct(n)ttgaagatca	cccaaggcct(c)ttgaagatca
167658	atttattttt(n)nnctctacatg	atttattttt(g)nnctctacatg
167659	tttattttin(n)ctctacatga	tttattttin(g)ctctacatga
186539	agctggagga(n)gnaaaacttc	agctggagga(g)gnaaaacttc
186541	ctggggaggg(n)aaaaactcta	ctggggaggg(g)aaaaactcta
196467	ttaaaattac(t)tgtaaaatct	ttaaaattac(t)tgtaaaatct
196468	taaaattacn(n)ntaaatctta	taaaattac(t)ntaaatctta



```
KEYWORDS      human.
SOURCE        Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 7260)
AUTHORS      Vernet,C.A., Fernandes,E., Shinkets,R.A., Macdougall,J. and
              Spaderna,S.K.
TITLE        Polyptides and nucleic acids encoding same
JOURNAL      Patent: WO 0162928-A 3 30-AUG-2001;
              Curagen Corporation (US)
FEATURES     Location/Qualifiers
              1..7260
                source          /organism="Homo sapiens"
                CDS             /db_xref="taxon:9606"
                                136..5211
                                /note="unnamed protein product"
                                /codon_start=1
                                /protein_id="CAC69812.1"
                                /db_xref="GI:15593890"
                                [translation="MASWTSPWMLIGVMHSPLPOTTAETKSPGAYFLPERALSPQG
                                SLEDTTGEFLTIRYDDQTSRNRSDKDNWDAGWSDCSRTCGGASVLRRC
                                LGRNCEGNIRYKTKSNHDCPDPAEDFRAQCSAYNDVQYGHYEWLPRYNDPAAP
                                CALKHAQGNLVVLAAPKLVLDGTRCNDSLDKMSIGICQAVGCDROLGSKNEDNG
                                VCAGDSTCLVRGQSKSHVPEKREENVIAVPLGRSVRTVKGPAHLFIESKTLQG
                                SKGEHFNPGVFEVNTTVEFGQSERQTFKIPGLMADFIPKTRYTAAKDSVVOFF
                                FYOPISHQNRQTDFFPCTVTCGGYOLNSAECVDIRLKVVPDHYHYYPENVKPKPK
                                LKESNDPCPSSDGRFKEIMPDHFQPLPWEHNPWTACSVSCGGGTOBRSFVCEVSM
                                HGEILQVWENKMYAPKPMQTCNLFDCPKWIAEWOSQCTVTCRGLRYRVVLCINH
                                RGEHVGCNPQLKHKIEBCVIPICYPKPKESVPBAKLPWLAKQALQDETRITREP
                                TIPEPWSASTTGPQSVQVREYKRVLLFTQTETELPEECEGKLETPRCPILLEA
                                CDEPASRELDIPPEDETSTWMEYAGFTPCATCGGHOEAIAVCLHIQTQVND
                                SLCDMVRPAMSAQCTEPCPRHVSWSGPCSATCGVGIOTRVDYVLPBGFTAPP
                                ECECRDKPHALQACNDFCPGWHIEWQOCSTCGSGTQNRVTCROLLTDGSEFLNL
                                SDELQGPKNASHKSCARTDCPHLAVGWSKCSVSCVGIQRRKQVQORLAKGRRI
                                PUSEMCRDLPGPLVRSQCMPECSKIKSEMTKLGEGQPIILSVQVRYIQTREERI
                                NUTIGSVRYLLPNTSVLIICPVRRFQKSLIOWKRGCLQNSKRLGITXSGSLKIHCL
                                AAPDIGVCIAGSAGTGVVLIQTDNRLIARPALREPMREYPGMDHSEANLGVTW
                                HKRMWNKNLDYLDDDHISNQPFLRALLGHSNAGSTNSWELKNQFEAAVKQGA
                                YSMDTAQDELRLNMSQLMETGEVSDLASOLYLVAELAKAQPTHMOWRGIOETP
                                PAAQLRGFTSVSQSSHAKNSGLTFKPKGPVLMRQSPSPISFNKTIINSRIQNTYVI
                                TKRTVINILCDTSEATYIWKDGLIQLPSVKIILDTGKIQLQNPTRKSGIYE
                                CSVNHQSDNDESVSSVLYAPVLIQSVERNITKPEHNLISVWGGVVEALGANVTIR
                                CPGVQPNQNIWMLKRGISLNSVLLFNGSLLOLQNSLENEGYVCIAIATNALKAVA
                                TSFVHLLERRWPESRIFVLQGHKKYILQATNRTNSNDPTGEPPQEPFEPWEPGNWHC
                                SATCGHLGARIORPQCMANGOEVSALCDHLQPLAGFEPENIRDCPARWFTSVWSQ
                                CSVSCGEGVHSROVTKRKTANGTVQVSPRACAPKDRPLGRKPCFSGHPCVQWENR
                                CPGRCMGRVAVQOORTACQHSNDSNCDRKRPTLRNCTSGACBVCWHTGPKWPKCT
                                AACGFGQSRKVDCHITRSCKPVAKRHCVOKKKPISWRHCLGPSCDRDCDITDTHYCMF
                                VKHLNLCJSLDRYKQRCQSQCEG*
BASE COUNT   2022 a 1593 c 1811 g 1834 t
ORIGIN
Query Match   19.3%; Score 1023.4; DB 6; Length 7260;
Best Local Similarity 60.9%; Pred. No. 2.3e-173;
Matches 1744; Conservative 0; Mismatches 1076; Indels 45; Gaps 3;
QY 60 cctgaggtccagaccacacgctccgagagagcggagcgccctatggaagcctg 119
Db 321 CCAGACCTCAAGAAACACTCTTCAGATGAAGACAAGATGGCAATGGGATCTTGGGG 380
QY 120 cccatggagtgaatgctcacgcacctcggggggtggggccctcctactctcgaggcgctg 179
Db 381 CGACTGGAGTGACTGCTCCCGACCTGTGGGGGAGGAGCATCATATCTCTCGGGAGATG 440
QY 180 cctgagcagcagagctgtgaaggaagaataatccgatatagacaacatgcagtaagtgtgga 239
Db 441 TTTGACTGGAGGAATTTGTAAGGGGACAGCAATTCGGGTACAAGACATGACGCAATCATGA 500
QY 240 ctgcccaccagaacagtgtaattccagctcagcaaatgtcagctcataatgatgtcaa 299
Db 501 CTGCCCTCCAGATGAAGAAATTTACAGAGCCCGACAGTGTCTCAGCCTACAAATGATGCCA 560
```



Qy	240	ctgccaccagaagcaggtgattccgagctcagcaatgctcagctcataatgagtcaa	299
Db	366	CTGCCCTCCAGATGCAGAGAAGATTTCAGAGCCAGCAGTGTCTCAGCCTACAATGATGTCCA	425
Qy	300	gcacatggccagtttatgaatggcttcctgtgtctaagtaccctgcacaacccaatgttc	359
Db	426	GTATCAGGGGCAATTACTATGAATGGCTTCCACGATATATGAATCATCTGTGTCGCCGGTGC	485
Qy	360	actcaagtgcgaagcgaaggacaacccctggttgtgaactagcacctaaagtctctaga	419
Db	486	ACTCAAGTGTATGCACRAGGACAAAATCTGCTGTGGCTGGCTGGCACCTAAGGTACTGGA	545
Qy	420	tggtaacgcttgctatacagaactctttggatatgtgcatacagtggtttatgccaattgt	479
Db	546	TGGAACCTGTTTGCACACCGACTCTCTGGACATGTGTATCAGTGGCATCTGTTCAGSCAT	605
Qy	480	tggctgcgactcacagctgggaagcacgctcaagaagaataactgtgggtctgcacaag	539
Db	606	GGCTGCCATCGGCACCTGGCAACTGGGAAGCAATGCCAAGAGGACAACTGTGGAGTCTGTGCCGG	665
Qy	540	agatgggtccacctcgcgctgggtccgagggcagttataaatcccagctctccgcaccaa	599
Db	666	CGATGGCTCCACCTCGAGCTGTACGGGGACAATCAAAAGTCACACGTTTCTCTCTGAAAA	725
Qy	600	atcggatgatactgtgttgaattccctatctgaaatgagacataattcgcttgtctiaaa	659
Db	726	AAGAAGAAATGTAAATGTCTGTCTTGTGGGAAGTCGAAGTGTGAGAATTACAGTGAA	785
Qy	660	aggtcctgatcactttatcttgaaaccaaaccctccaggggactaaagggtgaaacacag	719
Db	786	AGGACCTGCCACCCTTTATTGTAATCAAAAACACTTCAAGGAAGCAAGAGAGACACAG	845
Qy	720	tctcagctccacaggaaatttctctgtggacaattctagtgggacttccagaaatttcc	779
Db	846	CTTTAACAGCCCCGCGCTTTTGTGCTAGAAAACACACAGTGGAAATTCAGAGGGGGCTC	905
Qy	780	agacaaaagatactgagaatggctggaccactccacagcagatttcattgtcaagattcc	839
Db	906	CGAGAGGCAAACTTTTACATTTCCAGACCTCTGTATGGCTGTATTCATCTTCAAGACCAG	965
Qy	840	taactcgggtccgcgtgcagttacagttccagttccattcttctatcaacccatcatccacq	899
Db	966	GTACACTGCAGCCAAAGACAGCGTGGTTCAGTCTCTTTTACCAGCCCATCAGTCATCA	1025
Qy	900	atggaaggagacggaattcttctcgtgcagcaacctgtgcagaggttatacagctgac	959
Db	1026	GTGGAGACAAACTGACTCTTTTCCCTGCAGCTGTGACGTGTGGAGGAGTTATCAGCTCAA	1085
Qy	960	atcggctgagtgctacgatctcagagacaacgctgggtgctgaccaatactgtcacta	1019
Db	1086	TTCTGCTCAATGTGTTGATATCCGCTTGAAGAGGTAGTTTCTTGACCATTATTGTCACTA	1145
Qy	1020	ttaccagagaaacatacaacccaaacccaagcttcaggagtgcaacttggactcttctgtcc	1079
Db	1146	CTACCTCAAAATGTAAAACCAAAACCAAACTGAAGGAATGCAGCATGATCCCTGCC	1205
Qy	1080	agccagtgacgatatacagcagatcatgcttatgacctctaccatcccccttctcgggtg	1139
Db	1206	ATCAAGTATGATTTAAAGATAATAGCCCTATGACCATTCCCAACCTCTTCTCGCTG	1265
Qy	1140	ggagggccacccatggaccgctgtctctctcgtgtgggggggcatccagagccgggc	1199
Db	1266	GAAACATAATCTTGGATGTCATGTCCGTGTCTGTGGAGNGGATTCAGACACGGAG	1325
Qy	1200	agttctctgtgtggaggagacatccaggggcgatgctcacttcagtggaagagtggaaatg	1259
Db	1326	CTTTGTGTGTAGAGGAATCCATGCATGGAGAGATATTGCAGGTGGAAGAATGGAAGTG	1385
Qy	1260	catgtacaccccttaagatgcccatcgcagccctgcacaacttttgcctgcctaaatg	1319
Db	1386	CATGTACGCCCAACCAACCAAGGTATGCAAACTTCTAATCTCTTTGATGTGCCAAGTG	1445

Qy	1320	gctggccaggagtggtctccgtgcacagtcagatgtgtggccagggtccctcagataccctgtg	1379
Db	1446	GATTGCCATGGAGTGGTCTCAGTGCACAGTGACTGTCTGGCCGAGGGTTACGGTTTACGGGGT	1505
Qy	1380	ggctcctgcatacgacatcgaggaatgcacacagagagctctagcccaaaaaacaagaacc	1439
Db	1506	TGTTCTGTGATTAAACCACCGGAGAGCATGTTGGGGGCTGCAATCCCAACATGGAAGTT	1565
Qy	1440	ccacataaagaggaatgcatgtaccactccctgtctataaaccacaaagagaaacttcc	1499
Db	1566	ACACATCAAGAGAGATGTCTCATTCCTACCTCCGCTGTTATAACCAAAAGAAAAGTCC	1625
Qy	1500	agtcagagccaaattgcatgtttcatttcaaaacaaagctcaagagctagaagagagctgtgt	1559
Db	1626	AGTGAAGCAAAATTGCTTGGCTGAAACAAGCAAGAACATAGAAGAGACAGCAATAGC	1685
Qy	1560	gtcaagagagccctcgtttcatccagagagcctggctggcctgcacagtcacactgtgtgt	1619
Db	1686	AACAGAAGAACCAAGCTTCATCCAGAACCCCTGGTCAAGCTGACGTACACGTGTGGGCC	1745
Qy	1620	ggggaccacaggtgcgaatagtcaggtgcaggtgtcctgtcttcttctctcagtcctgtgc	1679
Db	1746	AGGTTGTCAGTCCGGAGGTGAAGTGCCTGTGCTTCCTACATTCACGCACATGAGAC	1805
Qy	1680	tgaacctgcctattgacagtgtaaggggtgaagggcccaagccagatccccagctgctgttatgc	1739
Db	1806	TGAGTCCCGGAGAGAGTGTGAAGGGCCCCAAGCTGCCACG-----GA	1850
Qy	1740	aggcccatgcagcggggaaattcctgagttcaaccacagacagacagatgggtctttgg	1799
Db	1851	ACGGCCTTGCTCTGTGAGACATGTGATGAGAGCCCGGCTTCCGAGAGCTTAGACATCCC	1910
Qy	1800	tggcctgcaggatttcgacgagctgtatgactgggagtagtaggggtttccacaagtgctc	1859
Db	1911	TCTCCCTCAGSACAGTCAGACGACTTACGACTGGGAGTACGCTGGGTTTACCCCTTGCAC	1970
Qy	1860	cgaactcctgtggaggaggtgtccagagagctgtgtgagctgcttgtaacaaacagactcg	1919
Db	1971	AGCAACATGCTTGGGGAGCCATCAAGAAAGCCATAGCAGTGTGTTTACATATCCAGACCA	2030
Qy	1920	ggagcctgctgaggagaacctgtgcctgaccagccgcgcgcacacagctcctgaagtc	1979
Db	2031	GCACACAGTCAATGACAGCTTGTTGATATGTTCCACCGCTCTCCAGCCATGAGCCAGGC	2090
Qy	1980	ctgcaatttgatccctgccagcaaggtgggaatttggcaagtgtagtccatgtagctct	2039
Db	2091	CTGTAAACACAGAGCCCTGTCCCCCGAGTGGCATGTGGGCTCTTGGGGCCCTGCTCAGC	2150
Qy	2040	cacatgtggggtgcgcclacagaccagagacgtctctctgcagccacctgcttccagaga	2099
Db	2151	TACCTGTGGATTGGAATTTCAGACCCGAGATGTACTGCCGTGACC-----	2197
Qy	2100	gatgaatgaacagtcactcctgctgatagctgtgtgcagcccaagccacgacacggt	2159
Db	2198	-----CAGGGGAGACCCTGCCCTCTCAGAGAGTCCGAGATGAAGACCCCATGCTTT	2252
Qy	2160	gcaagctgttaaccgctttaattgccccccagcctggttacctgcacagtgccagcgtg	2219
Db	2253	ACAAAGCATGCAATCAGTTTCTACTGCCCTCCTGGCTGGCACATTGAAGAATGSCACGATG	2312
Qy	2220	ttccagaacctgtggcgggggtgttcagaaaacgtgaggttttttgcagagcagcatggc	2279
Db	2313	TTCCAGGACTTGTGGCGGGGAACTCAGAACAGAGAGTCACTGTGCGGACGTGCTTAAC	2372
Qy	2280	tgatggcagcttctggagattcctgagaccttctgttcagcttcaaaacctgctgcaca	2339
Db	2373	GGATGGCAGCTTTTGGAAATCTCATGATGAATTTGCCAAGGACCCCAAGGCATGCTCA	2432
Qy	2340	gcaagcatgcaagaaatgactgtcccagcagtggtctctcagactggacagatg	2399
Db	2433	CAGTCTCTGTGCCAGGACACATGCTCTCCACATTTAGCTGTGGGAGACTGGTGGAGTG	2492
Qy	2400	ttccaagaactggcggggaagggcacccagactcgaagcgcatttgcgaagaatgcttaa	2459











RESULT<sup>9</sup>

AF237652 2523 bp mRNA linear PRI 02-MAR-2001  
 LOCUS Homo sapiens a disintegrin-like and metalloprotease domain with  
 DEFINITION thrombospondin type I motifs-like 3 (ADAMTSL3) mRNA, partial cds.  
 AF237652  
 ACCSSION AF237652  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 2523)  
 REFERENCE Hirohata, S., Anand-Apte, B., Seldin, M. and Apte, S.  
 AUTHORS Punctin, a member of a new family with similarities to ADAM-TS  
 TITLE proteases, is a component of extracellular matrix of skeletal  
 muscle  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2523)  
 AUTHORS Anand-Apte, B. and Apte, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-FEB-2000) Biomedical Engineering, Cleveland Clinic  
 Foundation, 9500 Euclid, Cleveland, OH 44195, USA  
 FEATURES  
 Location/Qualifiers  
 1..2523  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..>2523  
 /gene="ADAMTSL3"  
 225..>2523  
 /gene="ADAMTSL3"  
 /note="ADAMTSL3; similar to GenBank Accession Number  
 AB033059; extracellular matrix protein"  
 /codon\_start=1  
 /product="a disintegrin-like and metalloprotease domain  
 with thrombospondin type I motifs-like 3"  
 /protein\_id="AAK15041.1"  
 /db\_xref="GI:13183078"  
 /translation="MASWTSPWMLVIGVMFHSPLPQTAAKSPGAVFLPEFALSPQG  
 SLEDTGGQFITYRDDTTSRNRDDEKGNWDGWSDCSRFCGGGASVLRG  
 LTRNCEGQIRYKTSNHDQPDPAEDFAQQCSAYNDVQGYHYEWLPRYNDPAAP  
 CALKHAQGNLVVELAKVLGTRGNTSLDMCISGICQAVGCDKOLSNKEDNCG  
 VCAGDSTGLRVGQSKSVSPKREENVIAVPLGSRVRIIVKGPVHLFIESKTLQG  
 SKEHFSNFGVFPVETVQFSGSERQTEKIPLMADFIKTRYIAKDSVVOFF  
 FYQISHQRQDFFPTVTCGGYQILNSAECVDIRLKRVPDHYCHYTPENVKPKP  
 LKESMDPCSSDGEIIMPYDHFQPLRWEHNPWACSVSCGGGQIRSRFVCEESM  
 HGEILQVEWKKMYAPKPKVMOTCNLFDCPKWIAEWESQCTVTCGRGLRYRVLCNH  
 RGEHVGCCNPOLKHLIKRECVIPICYPKPKSPVEAKLPWLKQALEBETRIATEEP  
 TTRPEWSACSTTCGQVOVRVKRVLITFQTELPFEECEAGKPLETRICLLLEA  
 CDSPASREDILPDRSETTIDWEYAGFTPCATCLGGHGEAIVCLHITQQTVDN  
 SLCDMVHRPMSQACNTEPCPPRNHVGWSGPGSATCGVGIOTRDVYCLHPGETPAPP  
 EECRDEKPHALQACNQFDCCPGWHIP"  
 BASE COUNT 635 a 652 c 692 g 544 t  
 ORIGIN  
 Query Match 16.6%; Score 878.2; DB 9; Length 2523;  
 Best Local Similarity 65.1%; Pred. No. 2.5e-147;  
 Matches 1319; Conservative 0; Mismatches 693; Indels 15; Gaps 1;  
 QY 60 cctgaggtccagacgcacgtccgagagagagacggcgagcctatggatgcctggg 119  
 Db 410 CCAGACTCAAGAAACACTCGTTTCAGATGAAGCAAGATGGCACTGGGATGTTGGG 469  
 QY 120 cccatggagtgaatgctcacgcacctcgcggggtggggcctcctactctctgagcgctg 179  
 Db 470 CGACTGGAGTACTGTCTCCCGGACCTGTGGGGAGGAGGACATATATCTCTCGCGGAGATG 529  
 QY 180 cctgagcagcaagagctgtgaaggaagaataatccgatacacgaacatgcagtaagtggga 239  
 Db 530 TTGACTGGAAGGAATGTGAAGGGCAGAACTTCGGTACACAGCATGCAGCAATCATGA 589  
 QY 240 ctgcccaccagagaggtgatttcgagctcagcgaatgctcagctcgaatgatgtcaa 299

Db 590 CTGCGCTCCAGATGCAGAGATTTTCAGAGCCCGCAGAGTCTCAGCTACAATGATGTCCA 649  
 QY 300 gcacatggcaggtttatgaatggctctctgtgtctaatgacctgacacccatgttc 359  
 Db 650 GTATCAGGGGGCTTACTATGTAATGGCTTCCACGATATATATGATCTCTGTCGCCGTGTC 709  
 QY 360 actcaagtgcgaagcgaaggaaacccctgggtgttgaactagcacctaaagtgttaga 419  
 Db 710 ACTCAAGTGTCAATGCAAGGACAAACACTTGGTGTGGAGTGGCACCCTAAGGTACTGA 769  
 QY 420 tggtagcgtgtgtctatcacagaatctttgatatgcatcactggttattgccaattgt 479  
 Db 770 TGGAACTCGTTGCAACACGGACTCTTGGACATGATGATCAGTGGCATCTGTCAGCGAGT 829  
 QY 480 tggctgcgatcacacagctgggaagcacgcgtcaaggaataactgtgggtctgcaacgg 539  
 Db 830 GGGCTGGCATCGGCAACTGGGAAGCAATGCCAAGGAGGACAACTGTGGAGTCTGTGCCGG 889  
 QY 540 agatgggtccacctgcggtggtccgagggcagtgataataccccagctctccgcaccaa 599  
 Db 890 CGATGGCTCCACCTCGAGGCTGTACGGGGACAATCAAAGTCACACAGCTTCTCTGAAAA 949  
 QY 600 atcggtatgatactgtgtgcaattccctatggaagtagacatatccgctgtctttaa 659  
 Db 950 AAGAGAGAAAAATGTAATGTCTGTTCCTTGGGAAGTCGAAGTGTGAGAAATTCAGTGAA 1009  
 QY 660 aggtcctgatcacttatatcttgaaacaaacccctccagggagcaaatgaagtgaacag 719  
 Db 1010 AGGACCTGTCCACCTCTTTTATTAATCAAAACACTTCAAGSAAGCAAGGAGACACAG 1069  
 QY 720 tctcagctccacaggaacttccctgtggaacaattctagtgtggaactccagaattccc 779  
 Db 1070 CTTTAACAGCCCCGGCGCTTTGTGCTAGTAAACACAAACAGTGGAAATTCAGAGGGGCTC 1129  
 QY 780 agacaagagatactgagaatggctggaccactcacacagatcttcattctcaacccatccacgg 839  
 Db 1130 CGAGAGGCAAACTTTAAGATTCAGGACCTCTGATGGCTGATTTCACTTCAAGACCCAG 1189  
 QY 840 taactgggtcccgctgacagtagacagtagcagttctctatcaacccatccacccg 899  
 Db 1190 GTACACTGCAGCCAAAGACAGCGTGGTTTCAGTCTCTCTTTTACCAGCCCATCAGTCAICA 1249  
 QY 900 atgagagagacgattcttctctgctcagaacctgtgagagaggtttatcagctgac 959  
 Db 1250 GTGAGACAAACTGACTTCTTCCCTGCACTGTGACGCTGTGGAGGAGGTTTATCAGCTCAA 1309  
 QY 960 atcggtcgtggtcacgatctgagagagcaacccgtgtgtgtgtgtgacaaactactgtcacta 1019  
 Db 1310 TTCTGCTGAATGTGTGATATCCGCTTGAAGAGGGTAGTTCCTGACCATATATTGTCACTA 1369  
 QY 1020 ttaccagagaaacatacaacccaaacccaaactcaggagtgcaacttggaactctgtccc 1079  
 Db 1370 CTACCTGAAATGTAAACCCAAACAACTGAAGSAATGAGCATGGATGCTCCTGCC 1429  
 QY 1080 agccagtgcaggtacacagcagatcagcttctatgaccttaccatccctctcctggtg 1139  
 Db 1430 ATCAAGTGTGGATTTAAAGAGATAATGCCCTATGACCACTTCCAACTCTTCTCCTG 1489  
 QY 1140 ggaggccacccccctggaccgctgctcctcctcctgtgtgggggggcatccagacccgggc 1199  
 Db 1490 GGAACATAAATCTTGGACTGCAATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1549  
 QY 1200 agtttccgtgtgagagagacatccaggggcatctcactcagtggaagtggaaatg 1259  
 Db 1550 CTTTGTGTGTGTAGAGAAATCCATGTCATGGAGAGATATTGACGGTGAAGAATGGAAGTG 1609  
 QY 1260 catgtacacccctaaagtgcctcgcgcagccctgcaacatttttgaactgacctaaatg 1319  
 Db 1610 CATGTACGCCACCAACCCCAAGGTTATGCAAACTTGTAAATCTGTTTGTATTGCCCAAGTG 1669  
 QY 1320 gctgggcacagaggtggtctcctgacagtgacatgtggccagggtctcagatcacgtgt 1379



[illegible]

Db	2099	CTGTGGCGGGGAAACTCAGAACAGGAAGAGTCCACCTGTGGCAGCTGCTAAAGCGATGGCA	2158			
QY	2288	gtctcttgagcttctctgagacccttctgttcagcttcaaaacctgcctgccagcaagcat	2347			
Db	2159	GCTTTTGAATCTCTCAGATGAATTTGTGCCAAGGAGCCCAAGGCATCGTCTCACAAGTCCT	2218			
QY	2348	gcaagaagaatgactgtccagcgagtgcttctctctcagactggacagag	2397			
Db	2219	GTGCCAGGACAGACTGTCTCCACATTTAGCTGTGGGAGACTGGTTCGAAG	2268			
RESULT 11						
AX224815						
LOCUS	AX224815	2175 bp	DNA linear PAT 10-SEP-2001			
DEFINITION	Sequence 7 from Patent WO0161011.					
ACCESSION	AX224815					
VERSION	AX224815.1	GI:15554912				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 2175) Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B. and Sands,A.T.					
TITLE	Novel human thrombospondin repeat proteins and polynucleotides encoding the same					
JOURNAL	Patent: WO 0161011-A 7 23-AUG-2001; Lexicon Genetics Incorporated (US)					
FEATURES	Location/Qualifiers 1..2175 /organism="Homo sapiens" /db_xref="taxon:9606"					
BASE COUNT	583 a	532 c	575 g 485 t			
ORIGIN						
Query Match 16.08; Score 847.4; DB 6; Length 2175;						
Best Local Similarity 65.1%; Pred. No. 8.4e-142;						
Matches 1272; Conservative 0; Mismatches 666; Indels 15; Gaps 1;						
QY	60	cctgagttccaggaccgcacgtccctcgaggagaccggcgagcctatggatgcctgggg	119			
Db	186	CCAGACCTCAGAACAACACTCGTTGATGAGACAAAGATGGCAACTGGGATGCTGGGG	245			
QY	120	ccctggagtgaaatgctcaacgtcgggggtggggcctcctactctctgagggcgtg	179			
Db	246	CGACTGGAGTGACTGCTCCCGAGCTGTGGGGGAGGAGCATCATATTTCTCTCGGGAGATG	305			
QY	180	cctgagcagcagagcigtgaagaaataatccgatacagaaacatgcagtaaatgtgga	239			
Db	306	TTTGACTGGAAGGAATTTGTGAAGGGCAGACATTTGGGTACACAGCATGCGCAATCATGA	365			
QY	240	ctgcccaccagaagcaggtgatcttcagagctcagcaatgctcagctcataatgatgtcaa	299			
Db	366	CTGCCCTCCAGATGCAGAAGATTTCAGAGCCCAGCAGCTGCTAGCCTACATGATGTCCA	425			
QY	300	gcaccatggccagtttatgaatggcttctgtcttaatgacctgacaacccattctc	359			
Db	426	GTATCAGGGGCATTACTATGAATGGCTTCCACGATATAATGATCTCTGCTGCCCGTGTGC	485			
QY	360	actcaagtgcacaagcacaaggaacaacctggltgttgaaactagaccctaaagttttaga	419			
Db	486	ACTCAAGTGTATGCACAAGGACAAAACACTTGGTGGAGCTGGCAGCTAAAGTACTGGA	545			
QY	420	tggatcgcgttgcatacagaatctttggatatgtgcacagtgggtttatgccaaattgt	479			
Db	546	TGGAACCTCGTGTGAACACGGACTCCTTGGACATGTGATCAGTGGCATCTGTCAAGCAT	605			
QY	480	tggctgcgataccagctgggaagcaccgtcaaggagaataactgtgggggtctgcaacgg	539			
Db	606	GGGCTGGCATGGCACTGGGAGACATGCGAAGGAGCAACTGTGGAGTCTGTGCCGG	665			

QY	540	agatgggtccacctgcggctgctgcgagggcagatataaaatccacgtctctccgcacacaa	599
Db	666	CGATGGCTCCACCTGCAGGCTTATCGCGGCAACATCAAGTCAACGTTTCTCTCCGAAAA	725
QY	600	atcgatgatactggttgcaattccctatggaagttagacattatgcgctgtctttaa	659
Db	726	AGAGAAGAAATGTAATGTGTTCTTGGGAAGTCGAAGTGTGAGAAATTCACAGTGA	785
QY	660	aggtcttgatacttatctcggaaacccaaacccctccagggactaaaagtgtaaacag	719
Db	786	AGGACCTGCCACCCTTTATTGAATCAAAAACACTTCAGGAAGCAAAAGGAACACAG	845
QY	720	tctcagctccacagaaatttctctgtggacaattctctagtgtagcttccagaaatttcc	779
Db	846	CTTTTAACAGCCCGCGCTTTTGTCGTAGAAAACACACAGTGGAAATTCAGAGGGGCTC	905
QY	780	agacaagaagatactgagaatgtgtggaacctcaacagcagatttcaattgcaagattccg	839
Db	906	CGAGAGGCAAACTTTAAGATTCAGGACCTCTGATGGCTGATTTCATCTTCANGACCAG	965
QY	840	taactcgggtcgcgtgcacagtacagtcaggttcatctctctcaacccatcatccacog	899
Db	966	GTACACTGCAGCCAAGACACGGTGGTTCAGTTCCTTTTACCAGCCCATCAGTCATCA	1025
QY	900	atgaggggagacggatttcttctgtctcagcaacctgtggagaggttatcagctgac	959
Db	1026	GTGGAGACAACTGACTCTTTTCCCTGCACGTGCACGTGCAGCTGTGGAGAGGTTCATCAGTCAA	1085
QY	960	atcggtgagtgctacgattctgagagacaacccgtgtggtgtctgacaaatactgtcacta	1019
Db	1086	TTCTGCTGAATGTGGATATCCGCTTGAAGAGGGTAGTTCCTGACCATATTGTGCACATA	1145
QY	1020	ttaccagagaacatacaacccaaacccaagcttcagagtgagtgcaacttgatccttgcc	1079
Db	1146	CTACCTGAAATGTAAACCAAAACCAAACTGNAAGAAATGCAGCATGCCCTGCC	1205
QY	1080	agccagtgacggatacaagcagatcatgccttatgacctctaccatccccctctcogtg	1139
Db	1206	ATCAAGTGTATGGATTTAAAGAGATAATGCCCTATGACCACATTCACACTCTTCTCTCGCTG	1265
QY	1140	ggaggccaccccatggacccgctgctcctcctgtgtggggggggaatccagagccgggc	1199
Db	1266	GGAAACATAATCTTGACATGTCTCGCTGTCTCGTGGAGAGGGGAATCAGAGACGGAG	1325
QY	1200	agttcctgtgtgagagggacatccaggggccatgcaacttcagtggaagtggaatg	1259
Db	1326	CTTTTCTGTGTAGAGGAATCCATGCATGGAGAGATATTCGAGGTGGGAAGATGGAAGTG	1385
QY	1260	catgtacaccccttaagatgccatgcgcagccctgcaacatttttgaactgcctaaatg	1319
Db	1386	CATGTACGCCACCCAAACCCAAAGTTATGCAAACTGTGAATCTGTTTATTTGCCCAAGTG	1445
QY	1320	gctggacaggaatggtctccgtgcacagtgacatgtggccaggggcctcagataccgtgt	1379
Db	1446	GATTGCCATGGAGTGGTCTCAGTGCACAGTAGTCTGTGGCCGAGGGGTACGGTACCGGGT	1505
QY	1380	ggctcctgcatagcacatcagaggaatgcacacagagagctgtgagcccaaaaaaacagcc	1439
Db	1506	TGTTTCTGTATTAACACCCGGGACAGCATGTTGGGGGCHTGCAATCCACNACTGAAGTT	1565
QY	1440	ccacataaagaggaatgatcgtaccactccctgcgtataaaacccaaagaaacttcc	1499
Db	1566	ACACATCAAAAGAAATGTCTATTCCTCCATCCCGTGTATATAAACCAAAAAGAAAAAGTCC	1625
QY	1500	agtcagggcccaagtgccatggttcaaacaaagctcaagagctagaagaaggagctgctgt	1559
Db	1626	AGTGGAAAGCAAAAATTCGCTTGGCTGGAACAGCACAGAACTTGAAGAAGACAGCAATAGC	1685
QY	1560	gtcagagagccctcgttcatcccagagcctggtcgcctgcacagtcacactctgtgt	1619
Db	1686	AACAGAAGAACCAACCTTCATTCCAGAACCCCTGGTTCACCTGCGAGTACCAGCTGTGGCC	1745
QY	1620	qggqacccagtgccgaatgctcaqgtgcccaggtgctcctgtcttctctcagtcgctggc	1679

Db	1746	AGGTGTCAGGTCGCGAGGTGAAGTCCGCTGTCTCTCTACATTCACGCAGACTGAGAC	1805
Qy	1680	tgaacctgctattgacgagtgatgaaggcccaagccagcagcatcccagcgtgcctgttatgc	1739
Db	1806	TGAGCTGCCGAGGAAGAGTGTGAAGGCCGCCAAGCTGCCACC-----GA	1850
Qy	1740	aggcccaatgcagcgggaaattctctgagttcaaccagacagacagatgggtcctttgg	1799
Db	1851	ACGCCCTGCCCTCTGGAAGCATGTATGATGAGAGCGCCGCCCTCCCGAGAGTAGACATCCC	1910
Qy	1800	tggcctgcagatttcacgagctgtatgactggagtagtagaggggttcaccaagtctc	1859
Db	1911	TCCTCCCTGAGACAGTGTGAGACGACTTAGCAGCTGGAGTACGCTGGTTCACCCCTTGAC	1970
Qy	1860	cgagtctctgtaggagagtgctccaggagctgtgtgtagctgttgtagaacaaacagactcg	1919
Db	1971	AGCAACATGCTTGGGAGGCCATCAAGAGCCATAGCAGTGTGCTTACATATCCAGAGCCCA	2030
Qy	1920	ggagcctgctgagagaaacctgtgctgaccagccgcgcgcacacagctccttgaagtc	1979
Db	2031	GCAGACAGTCAATGACAGCTTGTGTGATATGGTCCACCGTCTCCAGCCATGAGCCAGGC	2090
Qy	1980	ctgcaatttgatccctgtcccagcaagtgggga	2012
Db	2091	CTGTAAACAGAGCCCTGTCCCCCCAGGAGAGA	2123
RESULT	12		
LOCUS	AX224825	8578 bp	DNA linear PAT 10-SEP-2001
DEFINITION	Sequence 17 from Patent WO0161011.		
ACCESSION	AX224825		
VERSION	AX224825.1	GI:15554917	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Donoho, G., Scoville, J., Turner, C.A., Friedrich, G., Zambrowicz, B. and Sands, A.T.		
TITLE	Novel human thrombospondin repeat proteins and polynucleotides encoding the same		
JOURNAL	Patent: WO 0161011-A 17 23-AUG-2001;		
FEATURES	Lexicon Genetics Incorporated (US)		
source	Location/Qualifiers		
	1..8578		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	2440 a 1827 c 2070 g 2241 t		
ORIGIN			
Query Match	15.8%;	Score 835.4;	DB 6; Length 8578;
Best Local Similarity	65.1%;	Pred. No. 1e-139;	
Matches 1272;	Conservative 0;	Mismatches 666;	Indels 17; Gaps 2;
Qy	60	ctctgagttccaggaacgcagcgtcccgagagagaccggagcgcctatgggatgcctgggg	119
Db	364	CCAGACCTCAAGAAACACTCGTTTCAGATGATGACAAAGATGGCAACTGGGATGCTTGGGG	423
Qy	120	cccatggagtgaatgactcaagcaacctcggggtggggctctactctctgaagcgtg	179
Db	424	CGACTGGAGTGAAGTGTCTCCGACCTGTGGGGGAGGAGCATCAATTCCTCGGAGATG	483
Qy	180	ctctgagcagcaagcgtgtgaagggaagaataatccgatcacagaacatgcagtaattgga	239
Db	484	TTTGACTGAAGAAATTGTGAGGGCCAGAACATTCGGTACAAAGATGCAGCAATCATGA	543
Qy	240	ctgcccaccagcagcaggtgtattccagagctcagcaatgctcagctcataatgatgtaa	299
Db	544	CTGCCCTCAGATGCAGAAAGATTTCAGAGCCCGACAGCTGCAGCTACAAATGATGTGCA	603

QY	300	gcacatggccagtttttatgaatggcttctctgtctaataagccctgcacaacccatgttc	359
DB	604	GTATCAGGGGCATTACTATCAATGGCTTCCACGATATATGATCTCGCTGCCCGCTGCC	663
QY	360	actcaagtgcacagccaaaaggaaacacccctggttgtgaactlagcacctaaaggtcttaga	419
DB	664	ACTCAAGTGTATGCACAGGACCAAACTTGTGTGGAGCTTGGCACTTAAGTACTGGA	723
QY	420	tggtcaecggttgctatacacgaactcttttgatattgcatatgcatcagttgtttatgccaaattgt	479
DB	724	TGGAACTCGTGTGCNAACGACGACTCTTGGACATGTGTATCAGTGGCATCTGTGACGCGAGT	783
QY	480	tggctgcgatcacacgtgggaagcaccctcaaggaagaataactctggggctcgcacacgg	539
DB	784	GGGCTGGCATCGGCACTGGGAAGCAATGCCAAGGAGGACAACTGTGGAGTCTGTGCCGG	843
QY	540	agatgggtccacactgccggctggctcgaaggcagtatataatcccagctctccgcacaacaa	599
DB	844	CGATGGCTCCACTCGACGGCTTGCAGGGGCAAACTCAAGTCAACAGCTTCTCTGAAAA	903
QY	600	atcggatgatactgtgttgcaattcccttatggaagttagacatatctgccttgtcttaaa	659
DB	904	AAGAGAGAAAATGTAAATGCTGTCTCTTTGGGAGTGAAGTGTGAGAAATTACAGTGAA	963
QY	560	aggtcctgatcacttatatatctgaaaccaaaccctccaggagactaaagtgaaacag	719
DB	964	AGGACCTGGCCACGCTCTTATTGAATCAAAAAACCTTCAAGGAAGCAAGGAGAACACAG	1023
QY	720	tctcagctccacaggaacttctctgtggacaattctagtgtggacttccagaattcc	779
DB	1024	CTTTAAACGCCCCGGCGTCTTTGTGTGATGAAAACACACAGTGAATTTTCAGAGGGGCTC	1083
QY	780	agacaaagagatactgagaatggctggaccactcacagcagatttctatgtccaagattcg	839
DB	1084	CGAGAGGCAAACTTTAAGATTCCAGGACCTCTGATGGCTGATTTCACTCTTCAGACCAG	1143
QY	840	taactcgggctccgctgacagtcacagtcacgttccattctctatcaaccatctccaccg	899
DB	1144	GTACACTGCAGCCAAAACAGACGGGTGGTTTCAGTTCTTCTTTACAGCCCATCAGTCATCA	1203
QY	900	atgagaggagacggattttcttctgtctcagcaacactgtgtggagaggtttatcagctgac	959
DB	1204	GTGGAGACAAACTGCATCTTTCCCTGCACTGTGACGTGTGGAGGAGGTTATCAAGCTCAA	1263
QY	960	atcggctgagctacgactgtgaggagcaaccggtgtgtctgaaccaactctacta	1019
DB	1264	TTCTGCTGAATGTGTGATATCCGCTTGAGAGGGTAGTTCTTGACCATTATTGTGCTACTA	1323
QY	1020	ttaccagagaaactcaaaccccaacccagcttcagagtgcaactggactgttccttgc	1079
DB	1324	CTACCTCGAAATGTAAAAACCAAAACCAAACTGAAGGAATGCAGCATGGATCCCTGCC	1383
QY	1080	agccagtgcagatatacagaacatcatgccttatgacctataccatcccccttctc--gg	1137
DB	1384	ATCAAGTGTGGATTAAAGAGATAATGCCCTATGACCACTTCCAAGCTCTTCTCTCGAGC	1443
QY	1138	tggagggccacccccatggaccgctgctcctcctcgtgtgggggggcattccagagccgg	1197
DB	1444	TGGGAACATAATCCCTTGGACTGCATGTTCCGTGTCTGTGGAGGAGGGATTCAGAGACGG	1503
QY	1198	gcagtttctctgtgtgagagagacatccaggggcgatgttcaacttcagtggagagtgga	1257
DB	1504	AGCTTTTGTGTGTAGAGGAATCCATGCATGGAGAGATATTGCAGGTGGAAGATGGAAG	1563
QY	1258	tgcattgaacccccctaagatgcccattcgcgagccgtgcaacatttttgactgccctaaa	1317
DB	1564	TGCATGTACGACCCCAACCCCAAGGTATTGCAAACTTGTAATCTGTTGATTGGCCCCAAG	1623
QY	1318	tggctggcacagagtggtctccgtgcacagtgcacatgtgtgcgaaggccctcagataccgt	1377
DB	1624	TGATTTGCCATTGGATGGTGGTCTCAGTGCACAGTGACTGTGGCCGAGGGTTACGGTACCG	1683

QY	1378	gtggtcctctgcacgaccatcgaggaatgcacacagggaggtgtgaccccaaaaacaaag	1437
Db	1684	gTGTGTTCTGTGTAATTAACCAACGCGAGAGCATGTTGGGGCTGCATCCACAACCTGAAG	1743
QY	1438	ccccacataaagagaaatgcctgtaaccactccctgctataaaccacaaagagaactt	1497
Db	1744	TTACACATCAAGAGAAATGTGTCTATCCCTCCCGTGTATAACCAAGAGAAAAGT	1803
QY	1498	ccagtcgagcccaagtgtccattggttcaaacaaagctcaagagctagaagagagtgct	1557
Db	1804	CCAGTGGAAAGCAAAATTCCTTGGCTGAACAAGCAACAAGACTAGAACAGACCAGAAATA	1863
QY	1558	gltgcaagagagccctcgctcaatccagagagccctggtcgccctgcaacagtcacctgtggt	1617
Db	1864	GCAACAGAAAGAACCAACGTTTCATTCGACAACCTGGTCAGCTGCAGTACCCAGTGTGGG	1923
QY	1618	gtggggaccaggtgcgaatagtcagggtgcaggtgtcctctgtcttctcagtcctgtg	1677
Db	1924	CCAGGTGTGCAGTCCGGGAGGTGAAGTGCCTGTGCTCTCACATTTCAGCGAGACTGAG	1983
QY	1678	gtgacctgcctattgaacagtgctgaagggtcccaagccagcatcccgagtgctgttat	1737
Db	1984	ACTGAGCTGCCCGAGGAAGAGTGTGAAGGCCCAAGCTGCCACC-----	2028
QY	1738	gcaggcccatgcagcgggaaattcctaattcaaccacagacagatgggtctctt	1797
Db	2029	GAACGGCCTGCCTCTGTGAAGCATGTGTATGAGAGCCCGGCTCCCGAGAGTAGACATC	2088
QY	1798	gdtggcctgcaggatttcgacagctgtatgactggagatgaggggttcaccaagtgc	1857
Db	2089	CCTCTCCCTGAGGACAGTGAGACGACTTACGACTGGGAGTACGCTGGGTTCACCCCTTGC	2148
QY	1858	tccgagtcctgtgagagggtgtccagagagctgtgtgtgagctgttttaacaaacagact	1917
Db	2149	ACAGCAACATGTTGGAGGCCATCAAGAAGCCAATGACAGTGTGTTTACATATCCAGACC	2208
QY	1918	eggagagcctgtgaggaagaacctgtgcgtgaccagccgcgcgcacacagctctgaag	1977
Db	2209	CAGCAGACAGTCAATGACAGCTTGTGTATGTGGTCCACCGTCTTCCAGCCATGACCCAG	2268
QY	1978	tcttgcatttggatccctgcccagcaaggtggga	2012
Db	2269	GCCTTAACACAGAGCCCTGTCCCCCAGGAGAGA	2303

RESULT#	13
LOCUS	AX224817
DEFINITION	Sequence 9 from Patent WO0161011.
ACCESSION	AX224817
VERSION	AX224817.1 GI:15554913
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Chordata; Mammalia; Eutheria; Primates; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1953)
REFERENCE	Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowsicz,B. and Sands,A.T. Novel human thrombospondin repeat proteins and polynucleotides encoding the same
AUTHORS	Patent: WO 0161011-A 9 23-AUG-2001;
TITLE	Lexicon Genetics Incorporated (US) Location/Qualifiers 1..1953
JOURNAL	/organism="Homo sapiens" /db_xref="taxon:9606"
FEATURES	
source	530 a 480 c 513 g 430 t
BASE COUNT	
ORIGIN	

Query Match 14.8%; Score 783.4; DB 6; Length 1953;  
Best Local Similarity 64.98; Pred. No. 2.4e-130;

	Matches 1184;	Conservative	0;	Mismatches 626;	Indels 15;	Gaps 1
QY	188	gcaagagctgtgaagggaagaataatccgatacagaacatcgagtaattgtgactgcccac	247			
Db	92	GAAGGAATTGTGAAGGCAGAACATTCCGTTACAAGACATGACGAATCATGACTGCCCTC	151			
QY	248	cagaagcaggtgatttccgagctcagcaatgtcagctcacaatgaatgatgtcaagcaccatg	307			
Db	152	CAGATGACAGAGATTTTCAGAGCCAGCAGTGTCTCAGGCTACAAATGATGTCCAGTATCAGG	211			
QY	308	gccagttttatgaatgcttctctgtctctaattgacctgacacccattgttcaactcaagt	367			
Db	212	GGCATTAATAATGAAATGGCTTCCACGATATAATGATCTCTGCTCCCGGTGTCACCTCAAGT	271			
QY	368	gccaagccaaagggaacacccctggttgtaactagcacctaaggtcttagatggtacgc	427			
Db	272	GTCTATGCAACAAGGACAAAACCTTGTTGTGGAGCTGGCACCTTAAGTACTGTGATGGAATC	331			
QY	428	gttgtctacagaatcttttgatgtgcatacagtggtttatgcacaaattgttggtctgc	487			
Db	332	GTTCACACAGGACTCTCTGGACATGTGTATCAGTGGCATCTGTTCAGCGATGGGCTGGC	391			
QY	488	atcaccagctgggaagcaccttcaagggaagataactgtgggtgtgcacaggaatgggt	547			
Db	392	ATCGGCAACTGGGAAGCAATCCCAAGGAGGACAACTGTGGAGTCTGTCCGCGGATGGCT	451			
QY	548	ccaactgcggctggttcgagggcagtaataatcccagctctccqcaaccaaaatcggatg	607			
Db	452	CCACCTCGAGGCTGTGTACGGGGACAATCAAAAGTACACAGCTTCTCTGAAAAAAGAGAAAG	511			
QY	608	atactgtgtgtgcaattccctatggaagtgcacataattgcgtctgtcttaaaaggtcctg	667			
Db	512	AAATATGTAATTGCTTCTCTTTGGGAAGTTCGAAGTGTGAGAATTACAGTGAAGGACCTG	571			
QY	668	atacttatatctggaaacccaaacccctccagggagactaaagggtgaacacagctctcagct	727			
Db	572	CCCACCTCTTTATTGAATCAAAACACATTCAGGAAGCAAAAGGAGAACACACAGCTTTAAACA	631			
QY	728	ccacagaacttctctgtgcaaatcttagtggacttccagaaattccagacaaaag	787			
Db	632	CCCCGGCGCTTTTGTCTGTAGNAACACACACAGTGGAAATTCAGAGGGGCTCCGAGAGGC	691			
QY	788	agatactgagatggctggacacccatccacagcagatttcaattgtcaagattogtaactcgg	847			
Db	692	AAACTTTTAAGATTCCAGAGCCCTGTATGGGTGATTTCATCTCAAGACACAGTACACTG	751			
QY	848	gtccgctgacagtacagttccagttcatcttctatcaaccatcatccacgaatgaggg	907			
Db	752	CAGCCAAAGACAGCGTGTTCAGTTCTTTTACAGCCCATCAGTCAATCAGTGGAGAC	811			
QY	908	agacgatttcttctctgtcagcaacctgtggaggaggttatacagctgacatcggtg	967			
Db	812	AAACTGACTCTTTCCCTGCATGTGTGCGTGTGGAGGAGTTATCAGCTCAATTCTGCTG	871			
QY	968	agtgctacgatctgaggagcaaccdgtgtgtgtgtgcaccaactactgtcaactattaccag	1027			
Db	872	AAATGTGTGATATCGCTTTGAAGAGGGTAGTTCCTGACCATTTATGTCACTACTACCTCG	931			
QY	1028	agaacatcaaacccaaaccccaagcttcaggagtgcaacttgatctctgtccagccagtg	1087			
Db	932	AAAAATGTAACACCAAAACCAAACTGAAGGAATGCAGCATGATCCCTGCCCATCAAGTG	991			
QY	1088	acggatacaagcagatcatgccttatgaacctctacatcccccttccctcggtggaggcca	1147			
Db	992	ATGGATTTAAAGAGATAATGCCTATGACCACTTCCACCTCTCTCTGCTGGACATA	1051			
QY	1148	cccacatgaaccgctgctctctctctcgtgtgggggggcatccagagccggcgagttcct	1207			
Db	1052	ATCCCTTGACTGCATGTTCCGTGCTCTGTGGAGGGGATTCAGAGACGGAGCTTTGTGT	1111			
QY	1208	gtgtggaggaggacatccaggggcatgtcacttcagtggaagatggaaatgcattgaca	1267			
Db	1112	GTGTAGAGGAATTCATGCAATGAGAGATATGTCAGTGTGGAAGAAATGGAAGTGCATGAC	1171			

QY	1268	cccttaagatgccatcgcgagccctgcaacatttttgactgccttaaatggctgacac	1327
Db	1172	CACCCAAACCCAAAGGTATGCAAACTTGTAATCTGTTTGATTTGCCCCAAAGTGGATGCCA	1331
QY	1328	aggagtgtctccgtgcacagtcagactgtggccagggcctcagatcacogtctgtctctct	1387
Db	1232	TGGAGTGTCTCAGTCACAGTACATTGTGGCGGAGGTTACGGTACCGGGTTCTTCTGTGT	1291
QY	1388	gcacgaccatcgaggaatgcacacagagggtgtagccccaaaaaagccccacataa	1447
Db	1292	GTATTAAACCCACCGGAGAGCACTTTGGGGGCTGCAATCCACAACATGAAGTTACACATCA	1351
QY	1448	aadaggaatgcactgtaccacactccctgctataaacccaaagagaacactccactcaggg	1507
Db	1352	ANGAAGATGTGTCATTCCCATCCCGTGTATTAAACCAAAAGAAAAGTCCAGTGGAG	1411
QY	1508	ccaagtgtccatggttcaaaagaagctcaagagctagaagaaggagctgtgtgccaagg	1567
Db	1412	CAAAATTTGCCCTTGGCTGAAACAAGACACAAGAACTAGAAGAGACCAGAATAGCAACAGAG	1471
QY	1568	agccctgcttcatccacagaggctgtggcctgcacagtcacctgtggtgtggggaccc	1627
Db	1472	AACCAAGCTTCATCCAGAACCCCTGCTCAGCCTGCAGTACCACGTGTGGGCCAGGTGTGC	1531
QY	1628	aggtgcgaatagtcgaagtgccaggggtcctgtcttctcagtcogtggctgacctgc	1687
Db	1532	AGGTCCGCGAGGTGAAGTGCCTGTGCTCCTCACATTACGCAGAGCTAGAGCTAGGTGC	1591
QY	1688	ctattgacagagtgaagggcccaagccagcatccagcgtgcctgtttatcagagcccat	1747
Db	1592	CGAGGAAGAGTGHGAGGCCCCCAAGCTGCCACC-----GAACGGCCCT	1636
QY	1748	gcagcggggaaattcctgagttcaacccacagacagacagatgggtctttgtgtggctgc	1807
Db	1637	GCCTCTTGGAAAGCATGTGATGAGAGCCCGGCCCTCCGAGAGCTAGACATTCCTCTCCCTG	1696
QY	1808	aggatttcgacagcttatgactggagtatgagggttccaaagtgcctccagtcct	1867
Db	1697	AGGACGTGAGACACTTACGATGGGAGTACGCTGGGTTTACCCCTTGCACAGCAACAT	1756
QY	1868	gtggagagggtgtccagagggctgtgtgagctgttgacaaacagacactcggagcctg	1927
Db	1757	GCTTTGGGAGGCCATCAAGAAGCCATAGCAGTGTGCTTTACATATCCAGACCCAGCAGACAG	1816
QY	1928	ctgaggagaacctgtgcgtgaccagccgcggccccccacacagctcctgaagtcctcaatt	1987
Db	1817	TCAATGACAGCTTGTGTGATATGGTTCACCGCTCTCCGCCATGAGCCAGGCTGTAAACA	1876
QY	1988	tggatccctgccacgaaggtggga	2012
Db	1877	CAGAGCCCTGTCCCCCCAGGAGAGA	1901
RESULT	14		
AX149471			
LOCUS	AX149471	2912 bp	DNA
DEFINITION	Sequence 27 from Patent WO0136638.	linear	PAT 08-JUN-

RESULT 14

AX149471  
LOCUS AX149471 2912 bp DNA linear PAT 08-JUN-2001  
DEFINITION Sequence 27 from Patent WO0136638.  
ACCESSION AX149471  
VERSION AX149471.1 GI:14347941  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 2912)  
Shimkets,R.A., Lichenstein,H., Vernet,C. and Fernandes,E.  
POLYPEPTIDES AND NUCLEIC ACIDS encoding same  
JOURNAL Parent: WO 0136638-A 27 25-MAY-2001;  
Curagen Corporation (US)  
FEATURES  
Location/Qualifiers  
1..2912  
/organism="Homo sapiens"

BASE COUNT	810 a	/db_xref="taxon:9606"	713 c	754 g	635 t
ORIGIN					
Query Match	12.8%	Score 676.8	DB 6	Length 2912	
Best Local Similarity	58.0%	Pred. No. 2.9e-111			
Matches 1281	Conservative 0	Mismatches 882	Indels 45	Gaps	
QY	717	cagctctcagctccacaggaacttccctgtgacaaattcagtgacagcttcaggaattccagaaatt	776		
Db	1	CAGCTTTAAACAGCCCCGGGCTCTTTGTCGTAGAAACACACAGTGGATTTTAGAGGGG	60		
QY	777	tcagacaaaagagatactgagaatggctggagaccactcacagcagatattcattgtcaagat	836		
Db	61	CTCGAGAGGCAAACTTTTAAAGATTCAGAGCCCTTTGATGGCTGATTTTCATCTTCAACAC	120		
QY	837	tcgtaactcgggctccgcgtgacagtlacagtcaggcttcattctctatcaaccatcatcca	896		
Db	121	CAGGTACACTGCACCAAGACAGCAGGTGGTTCAGTCTCTTTTACCAGCCCATCAGTCA	180		
QY	897	ccgatggaggagacggatttcttcctgtctcagcaacctgtggagaggttatcagct	956		
Db	181	TCAGTGGAGACAAACTGACTTCTTCCCTGCACTGTGACGTGTGGAGGAGGTTATCAGCT	240		
QY	957	gacatcggctgagtgctaacgatctgagagagcaacogtgggtgtctgcgaccaatctgtca	1016		
Db	241	CAATTCTGCTGAATGTGGATATCCGCTTGAGAGGGTAGTTCCTGACCACTATTGTGCA	300		
QY	1017	ctattaccagagaaacatacaaaccccaaaccccaagcttcaggagtgcaacttgagatcttg	1076		
Db	301	CTACTACCCTGAAATGTAAACCAACCAACCAACTGAAGAAATGCACATGGATGCCCTG	360		
QY	1077	tcacgcafdgcagcgatacaagcagatcatgccttatgacctctaccatccctctctctcg	1136		
Db	361	CCCATCAAGTGTGATTTAAAGAGATAATGCCCTATCACCACITCCAACCTCTTCTCTCG	420		
QY	1137	gtggagggcccccagagacogctgtcctcctctgtgtggggggggcatccagagccg	1196		
Db	421	CTGGGAACATAATCCTTGACATGCATGTTCGGTGTCCCTGTGGAGGAGGGATTCAGAGACG	480		
QY	1197	ggcagtttctctgtggaggagacatccaggggcagtcacttcagtggaagagtgaa	1256		
Db	481	GAGCTTGTGTGTAGAGGAATCCATGCGAGAGATATTGCAGGTGGAGAAATGSA	540		
QY	1257	atgcattgacaccctaaagatgcccatgcgcagccctgcacaatttttgactgccttaa	1316		
Db	541	GTGATGTACGACCCCAACCCAAAGTTATGCAAACTGTATCTGTTTGAITGCCCCAA	600		
QY	1317	atggctggcacaggagtggtctccqtcacagtgacatgtggccagggcctcagataccg	1376		
Db	601	GTGGATGCCATGGAGTGGTCTCAGTGCACAGTCACTTGTGCGCCAGGGTTACGGTACC	660		
QY	1377	tgtgtctctgcatacgaccatcagggaatgcacacaggagcgtgtagccccaaaaacaa	1436		
Db	661	GGTGTCTGTGTATTAAACCAACCGGAGAGCATGTTGGGGGCTGCAATCCACAACGTAA	720		
QY	1437	gccccacataaaggagaatgcattgcatcaccactcctctgtatatcccccaaaagaaact	1496		
Db	721	GTTACATCAAAAGAAAGATGTGCTATTCCTATCCCTGGTTTATAAACCAAAAGAAAAG	780		
QY	1497	tcacgtcaggcccaagtgtccattgttcaaacaaagctcaagagctagaagaaggagctgc	1556		
Db	781	TCCAGTGAAGCAAAATGGCTTGCTGAACACAGCACAGAACTACAGAGACCAGCAAT	840		
QY	1557	tgtgtcagaggagccctcgttcatccccagaggcctgtgtcgccctgcacagtcacctgtg	1616		
Db	841	AGCAACAGAAAGAACCAAGTTCATTCAGAAACCCCTGGTTCAGCCTGCAGTACCAGGTGG	900		
QY	1617	tgtgtgggaccaggtggcgaatagtcaggtggcagtgctcctgtcttctctcagtcctgt	1676		
Db	901	GCGGGGTGTCAGGTGCCGTGAGGTGAAGTGGCGTGTGCTCTCTCACTACAGCAGACTGA	960		



Db	2008	CGTGTGCTGCAGACTCCAAACGGCTTGGCATCACCAAGTCAGGCTACTATAAATCCA	2067
QY	2805	ccgctcaagccctcgatgcagcgcgtctacacgtctcagcggccggccggcgagca	2864
Db	2068	CGGTCTGCTGCCCGCACATCGCGGTGTCAGGCTCTGCACAGGAAC	2127
QY	2865	ctttgtattaaactcgcagcggcaaccgcaagctcgtggccggcc	2912
Db	2128	AGTTGTCTCAAGCTATTGTACTGTACAAACCGGCTCATCCAGGCC	2175
RESULT 15			
AX224811	AX224811	Sequence 3 from Patent WO0161011.	
LOCUS	AX224811	1341 bp	linear
DEFINITION	AX224811		
ACCESSION	AX224811.1	GI:15554910	
VERSION			
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 1341)		
	Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.		
	and Sands,A.T.		
TITLE	Novel human thombospondin repeat proteins and polynucleotides		
	encoding the same		
JOURNAL	Patent: WO 0161011-A 3 23-AUG-2001;		
FEATURES	Lexicon Genetics Incorporated (US)		
source	Location/Qualifiers		
	1..1341		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	357 a 315 c 349 g 320 t		
ORIGIN			
	Query Match	10.5%; Score 554.8; DB 6; Length 1341;	
	Best Local Similarity	68.1%; Pred. No. 2.1e-89;	
	Matches	787; Conservative 0; Mismatches 367; Indels 2; Gaps 1;	
QY	60	cctgagttccaggaccgcagctccgaggaggaccggagcgcctatgggatgctgggg	119
Db	186	CCAGACCTCAAGAACAACACTCGTTTCAGATGAGACAAAGATGGCAACTGGGTGCTGGG	245
QY	120	ccatgagtgaaatgcacgcacctcggggggtgggcccctactctctctgagcgctg	179
Db	246	CGACTGGAGTGACTGCTCCGGACCTGTGGGGGAGGAGCATCATATCTTCGCGGATG	305
QY	180	cctgagcagcaagagctgtgaaggagaataatccgatacagaacatgcagtaagtgtga	239
Db	306	TTTCACCTGGAGGAATTTGTAAGGGCGCAACATTCGGTACAAAGACATGCAGCAATCATGA	365
QY	240	ctgccaccagaacagtgatcttcagctcagcagatgcacactcagctataatgatgcaa	299
Db	366	CTGCCCTCCAGATGCAAGAAGATTCAGAGCCGACGAGTCACGCTACAATGATGTCCA	425
QY	300	gcaccatggccagttttatgaatggcttcctgtctaatgaccctgacaccccatgttc	359
Db	426	GTATCAGGGGCATTACTATGAATGSCITCCACGATATAATGATCCTGCTGCCCGTGTGC	485
QY	360	actcaagtgcacagccaaaggacaacccctgtgtgttgaaactagcacctaagggtttaga	419
Db	486	ACTCAAGTGTCTATGCACAAAGACAAAATCTGGTGGTGGAGCTGCGACCTAAAGGTACTGGA	545
QY	420	tggtagcgttgctatcacagaatctllggatgtgcatcagtggtttatgccaaattgt	479
Db	546	TGGAACTCGTTGCAACACCGGACTCCTTGGACATGTGTATCATAGTGGCATCTGTACAGCACT	605
QY	480	tggctgcgcatcacagctgggaagacacccgctcaaggaagataactgtgggttctgcacgg	539
Db	606	GGGCTGCGCATCGCAACTGGGAAGCAATGCGCAAGGAGACAACTGTGGAGTCTGTGCGCG	665

Qy	540	agatgggtccacgcgcgggctgggtccgagggcgagtagataaaatccacagctctccgcaaccaa	599
Db			
Db	666	CGATGGCTCCACCTGCAGCGCTTGTACGGGGAACAATCAAGTCACACGTTTCTCTCGAAAA	725
Qy	600	atcsgatgatactggttgcaattccctatggaatgagacatatccgcttgctcttaaa	659
Db			
Db	726	AAGAGAAGAAATGTATTGCTGTCTTCTTTGGGAAGTCGAAGTGTGAGAAATTACAGTGAA	785
Qy	660	aggtctctgataccttatatctgaaaccaaaccctccaggggactaaaggtgaaacag	719
Db			
Db	786	AGGACCTGCCCACTCTTTATTGAATCAAAAAACACTTCAAGGAAGCAAAAGGAGAACACAG	845
Qy	720	tctcagctccaaaggaacttccctgtggaacaattctagtgtggaattccagaataattcc	779
Db			
Db	846	CTTTAAACAGCCCGCGGCTCTTTGTGTGATAAACACACAGCTGGAAATTTACAGAGGGCGTC	905
Qy	780	agacaaagataactagagaatggctggaccactcacagacagatttcattgtcaagattcg	839
Db			
Db	906	CGAGAGGCCAAACTTTTAAAGATTCACGAGACCTCTGATGCTGANTTCATCTTCAAGACCAG	965
Qy	840	taactcgggctccgctgacagtagacagtcacagttccagttccattctctcaaacccatcatccacg	899
Db			
Db	966	GTACACTGACGCCAAAGACAGCGTGGTTCAGTTCCTCTTTTACACGCCATCAGCTCAICA	1025
Qy	900	atggagggagacggaattcttcttcctgtccagcaacctgtgagagaggttatcagctgac	959
Db			
Db	1026	GTGGAGACAAACTGACTTCTTCTCCCTGGCACCTGACGCTGGAGGAGGTTATCAGCTCAA	1085
Qy	960	atcggctgaatgctacgatactgaggagacaaccgtagtgggtgctgacaaactatgctcaacta	1019
Db			
Db	1086	TTCTGCTGAAATGTGTGATATCCGCTTGAGAGGGTAGTCTCTGACCATTATTGCTCACTA	1145
Qy	1020	ttaccagagaacatcaaacccaaacccagcttcagaggtgcaacttggaatcctgtcc	1079
Db			
Db	1146	CTACCTGAAATGTAAACCAAAACCAAACTGAGGAAATGCACATGGATCTCCTCGCCC	1205
Qy	1080	agccagtgcaggataacgacagatcatgccttatgacctaccatccccttcctc--gg	1137
Db			
Db	1206	ATCAAGTGTATGGATTTTAAAGAGATAATGCCCTATGACCACCTTCCAACCTCTTCGTCGAGC	1265
Qy	1138	tggagggccaccccatggaccggctcctcctcgtatgggggggcatccacagccgg	1197
Db			
Db	1266	TGGAGACATTAATCCTTGACCTGATGTTTCCGTGCTCTGTGGAGAGAGGATTCAGAGACGG	1325
Qy	1198	gcagtttccctgtgg	1213
Db			
Db	1326	AGCTTTGTGTGTGTAG	1341

Search completed: July 24, 2002, 02:20:30  
Job time: 13259 sec



Result No.	Score	Query Match	Length	DB ID	Description
1	5172.2	97.8	5720	22	AAF97891
2	5171.8	97.8	5769	22	ABA06468
3	3390.4	64.1	3559	22	AAF97926
4	2556.2	48.3	2689	21	AAA47458
5	2552.6	48.3	2689	21	AAA47482
6	1960.8	37.1	3045	21	AAZ98232
7	1912.8	36.2	2031	22	ABA06676
8	1573	29.7	1869	20	AAZ34146
9	1573	29.7	1869	21	AAC78532
					Human secreted pro
					Human cDNA SEQ ID
					Human secreted pro
					Human TANGO 224 co
					Human TANGO 224 co
					Human signal pep1
					Human cDNA SEQ ID
					Human PRO1071 nucl
					Human PRO1071 (CNO

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5172.2	97.8	5720	22	AAFY7891
2	5171.8	97.8	5769	22	ABA06468
3	3390.4	64.1	3559	22	AAF97926
4	2556.2	48.3	2689	21	AAA47458
5	2552.6	48.3	2689	21	AAA47482
6	1960.8	37.1	3045	21	AAZ98232
7	1912.8	36.2	2031	22	ABA06676
8	1573	29.7	1869	20	AAZ34146
9	1573	29.7	1869	21	AAC78532
					Human secreted pro
					Human cDNA SEQ ID
					Human secreted pro
					Human TANGO 224 co
					Human TANGO 224 co
					Human signal pep1
					Human cDNA SEQ ID
					Human PRO1071 nucl
					Human PRO1071 (CNO

PI Young PE, Wei P, Florence

XX WPI; 2001-235311/24.  
 XX Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. Cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 XX  
 PS Claim 1; Fig 4; 890pp; English.  
 XX  
 CC The present sequence encodes one of 32 novel human secreted polypeptides.  
 CC The nucleic acid molecules and polypeptides they encode may be used in  
 CC the prevention, diagnosis and treatment of diseases such as  
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus  
 CC and human immunodeficiency virus (HIV) infections), hyperproliferative  
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. sclerotic syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the  
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid  
 CC sequences in samples. The polypeptides may be used as antigens in the  
 CC production of antibodies and in assays to identify modulators of  
 CC their expression and activity.  
 XX  
 SQ Sequence 5720 BP; 1293 A; 1677 C; 1611 G; 1139 T; 0 other;

Query Match 97.8%; Score 5172.2; DB 22; Length 5720;  
 Best Local Similarity 99.0%; Pred. No. 0;  
 Matches 5235; Conservative 0; Mismatches 3; Indels 51; Gaps 1;

QY 1 atggaatgctgcctgcggaacatctcggcacactgctctcttcttggcttctcctc 60  
 DB 67 atggaatgctgcgctgcggaacatctcggcacactgctctcttcttggcttctcctc 126  
 QY 61 ctgagttccagaccgacgctccgagagagacggagcgcctatggatgctggggc 120  
 DB 127 ctgagttccagaccgacgctccgagagagacggagcgcctatggatgctggggc 186  
 QY 121 ccatggagtgaaatgctacgcacactgcggggggtggggcctctactctctgagcgctgc 180  
 DB 187 ccatggagtgaaatgctacgcacactgcggggggtggggcctctactctctgagcgctgc 246  
 QY 181 ctgagcagcagagctggaaggaagaaatccgatacagaacatgcagtaagtggac 240  
 DB 247 ctgagcagcagagctggaaggaagaaatccgatacagaacatgcagtaagtggac 306  
 QY 241 tggccacagaagcagtgatttccgagctcagcaatgctcagctcataatgatgcaag 300  
 DB 307 tggccacagaagcagtgatttccgagctcagcaatgctcagctcataatgatgcaag 366  
 QY 301 caccatggcagatttatgaatgctctcctgctgctaatgacctgacacccatgttca 360  
 DB 367 caccatggcagatttatgaatgctctcctgctgctaatgacctgacacccatgttca 426  
 QY 361 ctcaagtgcacaagcacaagacacccctggttggaaactagcacctaaaggtcttagat 420  
 DB 427 ctcaagtgcacaagcacaagacacccctggttggaaactagcacctaaaggtcttagat 486  
 QY 421 ggtacgctgtgctatacagaatcttggatattgctatcagtggtttatgccaattgtt 480  
 DB 487 ggtacgctgtgctatacagaatcttggatattgctatcagtggtttatgccaattgtt 546  
 QY 481 ggtcgcatacagctgggaagcacgcgtcaaggaataactgtgggtctgcaacgga 540  
 DB 547 ggtcgcatacagctgggaagcacgcgtcaaggaataactgtgggtctgcaacgga 606  
 QY 541 gatgggtccactcctgcgctggtccgagggcagtgataataatcccgactcctccgcaaccaa 600  
 DB 607 gatgggtccactcctgcgctggtccgagggcagtgataataatcccgactcctccgcaaccaa 666

QY 601 tcggatgatactgtggtgcaattccctatggaagtagacatatcgcttcttcttaaaa 660  
 DB 667 tcggatgatactgtggtgcaattccctatggaagtagacatatcgcttcttcttaaaa 726  
 QY 661 ggtcctgatacattatctggaacaaacccctccagggagtaaaagtgaaacagct 720  
 DB 727 ggtcctgatacattatctggaacaaacccctccagggagtaaaagtgaaacagct 786  
 QY 721 ctcaagctccacaggaacttctctgtggacaattctagtgtggacttccagaataattcca 780  
 DB 787 ctcaagctccacaggaacttctctgtggacaattctagtgtggacttccagaataattcca 846  
 QY 781 gacaaaagagatactgagaatggctggaccctcacagagatttcaattgtcaagattcgt 840  
 DB 847 gacaaaagagatactgagaatggctggaccctcacagagatttcaattgtcaagattcgt 906  
 QY 841 aactcggctccctgcctgacgtacagtcacagtcacagtcacattcttctatcaacccatcatcacoga 900  
 DB 907 aactcggctccctgcctgacgtacagtcacagtcacagtcacattcttctatcaacccatcatcacoga 966  
 QY 901 tggaggagagcaggtattcttcttctgctcagcaacccctgtggaggaggttatacagctgaca 960  
 DB 967 tggaggagagcaggtattcttcttctgctcagcaacccctgtggaggaggttatacagctgaca 1026  
 QY 961 tcggctgagtgctacgatactgagagacacccgtgtggtgtgtgacccaatactgtcaactat 1020  
 DB 1027 tcggctgagtgctacgatactgagagacacccgtgtggtgtgtgacccaatactgtcaactat 1086  
 QY 1021 taccagagaacatcaaaacccaaacccagcttcaggaagtcacacttgatcttctgtcca 1080  
 DB 1087 taccagagaacatcaaaacccaaacccagcttcaggaagtcacacttgatcttctgtcca 1146  
 QY 1081 gccagtgcgggatacagaacgacgacatgccttatgacctctaccatcccccttctcctcgtgg 1140  
 DB 1147 gccac-----ggtgg 1155  
 QY 1141 gagccaccccatggaccgcgtgctcctcctcgtgtggggggggtccatccagagcgggca 1200  
 DB 1156 gagccaccccatggaccgcgtgctcctcctcgtgtggggggggtccatccagagcgggca 1215  
 QY 1201 gtttctctgtggaggagacatccagggccttcacttcagtggaagagtggaatgc 1260  
 DB 1216 gtttctctgtggaggagacatccagggccttcacttcagtggaagagtggaatgc 1275  
 QY 1261 atgtacacccctaaagatgcccatcgccagcctgcaacatttttgactgcctctaaatgg 1320  
 DB 1276 atgtacacccctaaagatgcccatcgccagcctgcaacatttttgactgcctctaaatgg 1335  
 QY 1321 ctggcacaggagtggtctcctgacagtgacatgtgcccagggtccctcagataccgtgtg 1380  
 DB 1336 ctggcacaggagtggtctcctgacagtgacatgtgcccagggtccctcagataccgtgtg 1395  
 QY 1381 gtcctctgacatcgacatcgaggaaatgcacagagagctgtagcccaaaaaaagccc 1440  
 DB 1396 gtcctctgacatcgacatcgaggaaatgcacagagagctgtagcccaaaaaaagccc 1455  
 QY 1441 cacataaagaggaatgcatacctcctcctgataaccccaagagagaaactcca 1500  
 DB 1456 cacataaagaggaatgcatacctcctgataaccccaagagagaaactcca 1515  
 QY 1501 gtcagggccaaagtgtccatggttcaaacagctcaagagctagaagaaggagctctgtg 1560  
 DB 1516 gtcagggccaaagtgtccatggttcaaacagctcaagagctagaagaaggagctctgtg 1575  
 QY 1561 tcaagaggagcctcgttcatccacagagcctggtcgccctgcacagtcacactgtggtgtg 1620  
 DB 1576 tcaagaggagcctcgttcatcccaaaaggcctggtcgccctgcacagtcacactgtggtgtg 1635  
 QY 1621 gggacccaggtgcgaatagtgccaggtgcctcctgtctcttctctcagtcctgtggt 1680  
 DB 1636 gggacccaggtgcgaatagtgccaggtgcctcctgtctcttctctcagtcctgtggt 1695







[illegible]

QY	1201	gtttctctgtgtggagagagaaatccagaggggcaatgtcaacttaagtgtgaagagtggaatgc	1260
DB	1214	gtttctctgtgtggagagagaaatccagaggggcaatgtcaacttaagtgtgaagagtggaatgc	1273
QY	1261	atgtacacccctaagaatgccatcgcagcgcctgcaacatttttgactgccctaaatgg	1320
DB	1274	atgtacacccctaagaatgccatcgcagcgcctgcaacatttttgactgccctaaatgg	1333
QY	1321	ctggcacaggagtggtctccgtgacagtgacatgtgccaggggcctcagatacccgctgtg	1380
DB	1334	ctggcacaggagtggtctccgtgacagtgacatgtgccaggggcctcagatacccgctgtg	1393
QY	1381	gtcctctgcataccatcgaggaatgcacacagcagctgtagccccaaaaaacaagccc	1440
DB	1394	gtcctctgcataccatcgaggaatgcacacagcagctgtagccccaaaaaacaagccc	1453
QY	1441	cacataaaagaggaatgcatactaccactcctcgtataaaacccaaagagaaattcca	1500
DB	1454	cacataaaagaggaatgcatactaccactcctcgtataaaacccaaagagaaattcca	1513
QY	1501	gtcgaggccaagtgcacctgttcaaaaagctcaagagctagaagaaggactcgtgtg	1560
DB	1514	gtcgaggccaagtgtccattgttcaaaaagctcaagagctagaagaaggactcgtgtg	1573
QY	1561	tcagaggagcctcgtttcatccagagcctggtcgccgtcacagctcaactgtgtgtgtg	1620
DB	1574	tcagaggagcctcgtttcatccaragcctggtcgccgtcacagctcaactgtgtgtgtg	1633
QY	1621	gggacccagggtgcgaatagtcagggtgccagggtgctcctgtctttctcagtcctgtgct	1680
DB	1634	gggacccagggtgcgaatagtcagggtgccagggtgctcctgtctttctcagtcctgtgct	1693
QY	1681	gacctgcctatgacagatgtgaagggcccaagcagcatccagcgtgccctgttatgca	1740
DB	1694	gacctgcctatgacagatgtgaagggcccaagcagcatccagcgtgccctgttatgca	1753
QY	1741	ggcccatgcagcgggaaattcctgattcaacccacgacgacagatgggctcttttgt	1800
DB	1754	ggcccatgcagcgggaaattcctgattcaacccacgacgacagatgggctcttttgt	1813
QY	1801	ggcctcaggaatttcacagagctgtatgactgggaatgtagggggttccccaagtctcc	1860
DB	1814	ggcctcaggaatttcacagagctgtatgactgggaatgtagggggttccccaagtctcc	1873
QY	1861	gagtcctgtggaggaggtgtccaggaggtgtgtgtgagctgtgtgacaacaacagactcg	1920
DB	1874	gagtcctgtggaggaggtgtccaggaggtgtgtgtgagctgtgtgacaacaacagactcg	1933
QY	1921	gagcctgtctgaggagaaacctgtcgtgaccagccgcgcggccccacagctcctgaagtc	1980
DB	1934	gagcctgtctgaggagaaacctgtcgtgaccagccgcgcggccccacagctcctgaagtc	1993
QY	1981	tgcgaatttgatccctgccacagaagtgggaattggcaagtggagtcgcatgtagtcctc	2040
DB	1994	tgcgaatttgatccctgccacagaagtgggaattggcaagtggagtcgcatgtagtcctc	2053
QY	2041	acatgtgggctcgccctacagacagacagctctctcgaccacacctgtttccagagag	2100
DB	2054	acatgtgggctcgccctacagacagacagctctctcgaccacacctgtttccagagag	2113
QY	2101	atgaatgaacaagtatccttggtgatgactgtgtgcagcccgaagccacacagtg	2160
DB	2114	atgaatgaacaagtatccttggtgatgactgtgtgcagcccgaagccacacagtg	2173
QY	2161	caagcttgtaacccgtttaattgcccccagcctgttaccctgcacagtgggcagccgtgt	2220
DB	2174	caagcttgtaacccgtttaattgcccccagcctgttaccctgcacagtgggcagccgtgt	2233
QY	2221	tcacagaacgtgtgagcggggtgttcagaacaagttaggtttctttgcaagcagcctagct	2280
DB	2234	tcacagaacgtgtgagcggggtgttcagaacaagttaggtttctttgcaagcagcctagct	2293
QY	2281	gatgcagactcctggagcttctggaacctctggaacctctgttcagcttcaaaacctcctgcag	2340

|||||  
Db 2294 gatggcagcttctgagcttctgagacattctgttcagcttcaaaaactgctgccag 2353  
QY 2341 caagcatgaagaaagatgactgttccacgagtgcttctctcagactggacagagtg 2400  
Db 2354 caagcatgaagaaagatgactgttccacgagtgcttctctcagactggacagagtg 2413  
QY 2401 tcacaaagctgcggggaagcaccagactcgaagcgccatttgcgaagatgctgaaa 2460  
Db 2414 tcacaaagctgcggggaagcaccagactcgaagcgccatttgcgaagatgctgaaa 2473  
QY 2461 accgacctcaccagctgttcacattccaccctgtgcgcgcctgtcttctctctcc 2520  
Db 2474 accgacctcaccagctgttcacattccaccctgtgcgcgcctgtcttctctctcc 2533  
QY 2521 atcaggccctgtatctggcaacctgtgcaaggcccgcgcccatccacgaagcacag 2580  
Db 2534 atcaggccctgtatctggcaacctgtgcaaggcccgcgcccatccacgaagcacag 2593  
QY 2581 ccgcacatcgccgcgcagagaaggtatcatccagactcgagcgaggaagtgcaac 2640  
Db 2594 ccgcacatcgccgcgcagagaaggtatcatccagactcgagcgaggaagtgcaac 2653  
QY 2641 tctgtgtgggggttcgctacactgtccccaagacggcggtgtgtgtgtgtgtgtgt 2700  
Db 2654 tctgtgtgggggttcgctacactgtccccaagacggcggtgtgtgtgtgtgtgtgt 2713  
QY 2701 gcgcagaggttcgcaagccctcatcactggagaaggcggcgacacctcatcagc 2760  
Db 2714 gcgcagaggttcgcaagccctcatcactggagaaggcggcgacacctcatcagc 2773  
QY 2761 tcgacgacgttcaacggtgtgcccccttcggtatctcagatccaccgctcaagccctcg 2820  
Db 2774 tcgacgacgttcaacggtgtgcccccttcggtatctcagatccaccgctcaagccctcg 2833  
QY 2821 gatcagcggtctacactgtcagcgggcgccggccggtggagcacttctgtatcagctc 2880  
Db 2834 gatcagcggtctacactgtcagcgggcgccggccggtggagcacttctgtatcagctc 2893  
QY 2881 atcggaggaacgcgaagctgtgcccgccttgagccgagaaatgaggaagaggtg 2940  
Db 2894 atcggaggaacgcgaagctgtgcccgccttgagccgagaaatgaggaagaggtg 2953  
QY 2941 cttgcggggaggaaggcgccgaagagagccctgcagacccaacaacacgaacaggg 3000  
Db 2954 cttgcggggaggaaggcgccgaagagagccctgcagacccaacaacacgaacaggg 3013  
QY 3001 atcttccaaacggcgaaggcggaagcgggcgctggccgcgaacccgggagccgc 3060  
Db 3014 atcttccaaacggcgaaggcggaagcgggcgctggccgcgaacccgggagccgc 3073  
QY 3061 tacgacgacctgtctcccgctgtctggaacagggcggtggccggagagctgtgccc 3120  
Db 3074 tacgacgacctgtctcccgctgtctggaacagggcggtggccggagagctgtgccc 3133  
QY 3121 tctgtggagcgaggaactcgcgcggaaggaacacagactcgaagaggaacccgggtgca 3180  
Db 3134 tctgtggagcgaggaactcgcgcggaaggaacacagactcgaagaggaacccgggtgca 3193  
QY 3181 gagcaagtctctcgtcactgtcccttcaaatggtgacgagcagcgcgccctggagcag 3240  
Db 3194 gagcaagtctctcgtcactgtcccttcaaatggtgacgagcagcgcgccctggagcag 3253  
QY 3241 atcttggaacactctccagacagccgaggaagctgcgcgacctctacagcaagcactg 3300  
Db 3254 atcttggaacactctccagacagccgaggaagctgcgcgacctctacagcaagcactg 3313  
QY 3301 gtggccagctgtggccagagatcttcccgagccacctgagacacagacagcgtctcg 3360  
Db 3314 gtggccagctgtggccagagatcttcccgagccacctgagacacagacagcgtctcg 3373  
QY 3361 aagccctcgagcgaggacttcccagtgactctctcctcataaacaacgtgtctggc 3420  
|||||

Db 3374 aagccctcgagcgaggaacttcccagtgactctctcgctcataaacacgtgtctggc 3433  
QY 3421 ttacagagcttccctcgaggaacttccacgggagcgcggggagagctctcgaagccca 3480  
Db 3434 ttacagagcttccctcgaggaacttccacgggagcgcggggagagctctcgaagccca 3493  
QY 3481 caccgaagcccaacatctgcgcaagatctcagcgccagcagctctcagcctcgag 3540  
Db 3494 caccgaagcccaacatctgcgcaagatctcagcgccagcagctctcagcctcgag 3553  
QY 3541 gtggtcaccacactggggcagacgggtggcccttgccagcgaggaactgagtctctcg 3600  
Db 3554 gtggtcaccacactggggcagacgggtggcccttgccagcgaggaactgagtctctcg 3613  
QY 3601 cactgtaggccatcgccaccacagggcctacacagctggggcccggaatggagaagaa 3660  
Db 3614 cactgtaggccatcgccaccacagggcctacacagctggggcccggaatggagaagaa 3673  
QY 3661 gtctagttcagtgacaggaattctctacagcagatgattctacagatcttggacca 3720  
Db 3674 gtctagttcagtgacaggaattctctacagcagatgattctacagatcttggacca 3733  
QY 3721 gtggaagcagatgtggtttctacacttgcaatggccaatgccttgagatacagctct 3780  
Db 3734 gtggaagcagatgtggtttctacacttgcaatggccaatgccttgagatacagctct 3793  
QY 3781 gtctccattgcctcacattagcgaagccactagtgaacacgtcacgaatgacagt 3840  
Db 3794 gtctccattgcctcacattagcgaagccactagtgaacacgtcacgaatgacagt 3853  
QY 3841 atcaacacggagaagcctgcagtcacagtcagataggaagcaccatcaaacagtgca 3900  
Db 3854 atcaacacggagaagcctgcagtcacagtcagataggaagcaccatcaaacagtgca 3913  
QY 3901 ggaagtgaatgtgacaatcaactggcaggtgaggaagtgcttgcagagctcttgcctc 3960  
Db 3914 ggaagtgaatgtgacaatcaactggcaggtgaggaagtgcttgcagagctcttgcctc 3973  
QY 3961 ttacaggaataaaacaaactgggtcccgccacatctgcacgaaggtctcttgcctc 4020  
Db 3974 ttacaggaataaaacaaactgggtcccgccacatctgcacgaaggtctcttgcctc 4033  
QY 4021 acaaacgtgtcctcctcggaacagggcctgactcctgcagggcgcccaatcttctatgga 4080  
Db 4034 acaaacgtgtcctcctcggaacagggcctgactcctgcagggcgcccaatcttctatgga 4093  
QY 4081 gagctgactgagacacccagctgctgactagatcccccccaagtcaccaacaggtg 4140  
Db 4094 gagctgactgagacacccagctgctgactagatcccccccaagtcaccaacaggtg 4153  
QY 4141 gaagacatcaggcgcttgcctgcctggcactggacgaaccttctcagtgactgactct 4200  
Db 4154 gaagacatcaggcgcttgcctgcctggcactggacgaaccttctcagtgactgactct 4213  
QY 4201 cctctgggaacacagctgttctgcctggatcctgggaattctctccttgcctgcccac 4260  
Db 4214 cctctgggaacacagctgttctgcctggatcctgggaattctctccttgcctgcccac 4273  
QY 4261 aaagggtcaccctgtccctaataatcactgggttctcagtggtgagccaattgtcactg 4320  
Db 4274 aaagggtcaccctgtccctaataatcactgggttctcagtggtgagccaattgtcactg 4333  
QY 4321 acaggactgacgacatcacatcttggcagctggacagatccttcaagtgcacaccttagc 4380  
Db 4334 acaggactgacgacatcacatcttggcagctggacagatccttcaagtgcacaccttagc 4393  
QY 4381 ggtgggtctcaagggaatttcagctgccttgcctgcagaaatgagcgaggggtgctcatgca 4440  
Db 4394 ggtgggtctcaagggaatttcagctgccttgcctgcagaaatgagcgaggggtgctcatgca 4453  
QY 4441 aaggcatctttagtgcacaaagattactggtgtgtgtgagacagctggcaacactgca 4500  
Db 4454 aaggcatctttagtgcacaaagattactggtgtgtgtgagacagctggcaacactgca 4513



protein coordinate data; infection; ss.

Homo sapiens.  
WO200121658-A1.

29-MAR-2001.

22-SEP-2000; 2000WO-US26013.

24-SEP-1999; 99US-0155709.

(HUMA-) HUMAN GENOME SCI INC.

Ni J, Baker KP, Hirse CB, Ebner R, Fiscella M, Komatsoulis GA; Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR; Young PE, Wei P, Florence KA;

WPI; 2001-235311/24.

Nucleic acids encoding 32 human secreted polypeptides, useful for preventing, diagnosing and/or treating e.g. cancers, Parkinson's disease and diabetic retinopathy -

Claim 1; Page 747-748; 890pp; English.

The present sequence encodes one of 32 novel human secreted polypeptides. The nucleic acid molecules and polypeptides they encode may be used in the prevention, diagnosis and treatment of diseases such as immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus and human immunodeficiency virus (HIV) infections), hyperproliferative disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases (e.g. Schmitzer syndrome, Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic disorders (e.g. corneal graft neovascularisation and diabetic retinopathy), neurological disorders (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease), infectious diseases and/or for promoting wound healing, regeneration and/or chemotaxis. The nucleic acid molecules may be used to produce the secreted polypeptides. They may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples. The polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators of their expression and activity.

Sequence 3559 BP; 807 A; 1037 C; 1005 G; 698 T; 12 other;

Query Match 64.1%; Score 3390.4; DB 22; Length 3559;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 3510; Conservative 11; Mismatches 18; Indels 56; Gaps 6;

QY 222 aacatgcagtaagtgtgagctgcccacgaagaagcaggtgatttcgagctcagcaatgctc 281

Db 18 aacatgcagctcttggaaatgccacacaaagcmsgtgalttccgagctcagcaatgctc 77

QY 282 agctcataatgatgtcaagcaccatgcccagtttatgaatgcttctgtctaatga 341

Db 78 agctcataatgatgtcaagcaccatgcccagtttatgaatgcttctgtctaatga 137

QY 342 cccctgacacccatgttcaactcaagtgcacgaagcacaacccctggtgttgaaact 401

Db 138 cccctgacacccatgttcaactcaagtgcacgaagcacaacccctggtgttgaaact 197

QY 402 agcacctaaagttcttagatggtacgcggttgcctatcacagaattcttgatgtgcatcag 461

Db 198 agcacctaaagttcttagatggtacgcggttgcctatcacagaattcttgatgtgcatcag 257

QY 462 tggtttatgcacaaattgttgctgcgcatccagctggaagcaccctgcaagaagataa 521

Db 258 tggtttatgcacaaattgttgctgcgcatccagctggaagcaccctgcaagaagataa 317

QY 522 ctgtggggtctgcaacgagatgggtccacctgcgctggtccgagggcagataaatc 581

Db 522 ctgtggggtctgcaacgagatgggtccacctgcgctggtccgagggcagataaatc 581

QY 4501 gctcctgtgttaacgggggttcagcagcccgcttaagtgctgctgaacagcag 4560

Db 4514 gctcctgtgttaacgggggttcagcagcccgcttaagtgctgctgaacagcag 4573

QY 4561 gaggtcaacccctgcccactgcaggaaggttcgcccctgcggtgcagccatcgctgc 4620

Db 4574 gaggtcaacccctgcccactgcaggaaggttcgcccctgcggtgcagccatcgctgc 4633

QY 4621 aaccgagagactgccttctcgtgtggtgagctcctggtgacctcctgctgctgaacccgagc 4680

Db 4634 aaccgagagactgccttctcgtgtggtgagctcctggtgacctcctgctgctgaacccgagc 4693

QY 4681 tgtggggagagtgctcagaccccgaggtgacctgtcaaaagctgaaagcctctgggagtc 4740

Db 4694 tgtggggagagtgctcagaccccgaggtgacctgtcaaaagctgaaagcctctgggagtc 4753

QY 4741 tccacccctgtgccatgacatgtgcacccaggtgcacagcgccgtgtggacacccag 4800

Db 4754 tccacccctgtgccatgacatgtgcacccaggtgcacagcgccgtgtggacacccag 4813

QY 4801 gctgttaaccagcagctgtgtgaggtgggccccttccagctggtggccagtgcaatggg 4860

Db 4814 gctgttaaccagcagctgtgtgaggtgggccccttccagctggtggccagtgcaatggg 4873

QY 4861 ccttgcatcgggctcactagctgtgcaacacagacaaagtcttctccagacacggagat 4920

Db 4874 ccttgcatcgggctcactagctgtgcaacacagacaaagtcttctccagacacggagat 4933

QY 4921 ggcacaccttaccatcagagcagtgagtgctcttccagagccctgtgagcaccagaac 4980

Db 4934 ggcacaccttaccatcagagcagtgagtgctcttccagagccctgtgagcaccagaac 4993

QY 4981 tgcgtggtcagagggcctgcagtgtaacctggagagtgagcctgtggacccctgtgcacagt 5040

Db 4994 tgcgtggtcagagggcctgcagtgtaacctggagagtgagcctgtggacccctgtgcacagt 5053

QY 5041 acctgtggaactcagctccagctccgcgtgtggaggtgtgtcatgcccgaccacac 5100

Db 5054 acctgtggaactcagctccagctccgcgtgtggaggtgtgtcatgcccgaccacac 5113

QY 5101 aagcgagtgctcagcagcctgtgctcctggggggcccccgcctgcacactgacgcgtgc 5160

Db 5114 aagcgagtgctcagcagcctgtgctcctggggggcccccgcctgcacactgacgcgtgc 5173

QY 5161 aacatcaccatgtgaaacatggagtgagagacacacacacagctactgcgagaggtg 5220

Db 5174 aacatcaccatgtgaaacatggagtgagagacacacacacagctactgcgagaggtg 5233

QY 5221 aacacgtgaaactctgcaactcagccagtttaaatctcgtcgtgtggaactgtggc 5280

Db 5234 aacacgtgaaactctgcaactcagccagtttaaatctcgtcgtgtggaactgtggc 5293

QY 5281 aaagcgtga 5289

Db 5294 aaagcgtga 5302

RESULT 3

AAF97926

ID AAF97926 standard; cDNA; 3559 BP.

XX AC AAF97926;

XX DT 01-JUN-2001 (first entry)

XX DE Human secreted protein cDNA, SEQ ID NO: 53.

XX DE Human; secreted protein; immunomodulatory; antisclerotic;

KW dermatological; anti-inflammatory; anti-HIV; cytostatic; cardiant;

KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;

KW nontropic; anticonvulsant; antialzheimer's; antiparkinsonian;

KW antimicrobial; vulnerrary; vaccine; gene therapy; cancer;



Db 318 ctgtgggtctgtgaacggagatgggtccacctgcggcgtgtccggaggccagtgataaactc 377  
QY 582 ccagctctcgcgaacaaatcgatgatactgtgtgtgcaattccctatggaagttagaca 641  
Db 378 ccagctctcgcgaacaaatcgatgatactgtgtgtgcaattccctatggaagttagaca 437  
QY 642 tattcgctgtcttaaaagttcgtgatcaactatatacttgaaaacccaaacccctccagg 701  
Db 438 tattcgctgtcttaaaagttcgtgatcaactatatacttgaaaacccaaacccctccagg 497  
QY 702 gactaaaggtgaaacagttcagctccacaggaaacttctctgtgacaaattctagtgt 761  
Db 498 gactaaaggtgaaacagttcagctccacaggaaacttctctgtgacaaattctagtgt 557  
QY 762 gacttccagaataattccagacaaagagatactgagaaatggctggaccactccacagaga 821  
Db 558 ggaacttccagaataattccagacaaagagatactgagaaatggctggaccactccacagaga 617  
QY 822 ttctcattgtcaagattcgttaactcgggctccgctgacagtacagttcagttctctcta 881  
Db 618 ttctcattgtcaagattcgttaactcgggctccgctgacagtacagttcagttctctcta 677  
QY 882 tcaacctcatccacogattggaggagacggatttcttctctgtcagcaacctgtgg 941  
Db 678 tcaacctcatccacogattggaggagacggatttcttctctgtcagcaacctgtgg 737  
QY 942 agagatttatcagctgacatcgcgtgagtgactagatctgagagcaacctgtgtgttc 1001  
Db 738 agagatttatcagctgacatcgcgtgagtgactagatctgagagcaacctgtgtgttc 797  
QY 1002 tgaccaatactgtcactattaccacagaagaacatacaacccaaacccaaagttcaggagtg 1061  
Db 798 tgaccaatactgtcactattaccacagaagaacatacaacccaaacccaaagttcaggagtg 857  
QY 1062 caacttggatcctgttccagccagtgacggatatacaagcagatcatgctctatgacctcta 1121  
Db 858 caacttggatcctgttccagcca----- 880  
QY 1122 ccatacccttctcgtgtggaggccaccccatggaacoggtgtctctcctcgtgtgggg 1181  
Db 881 -----ggtggaggccaccccatggaacoggtgtctctcctcgtgtgggg 926  
QY 1182 gggcatccagagcgggagtttctgtgtgagaggagacatccaggggcatctcacttc 1241  
Db 927 gggcatccagagcgggagtttctgtgtgagaggagacatccaggggcatctcacttc 986  
QY 1242 agtggaaagtggaaatgcattgacacccctaaagtgcacatcgccatcgccagccctgcaacat 1301  
Db 987 agtggaaagtggaaatgcattgacacccctaaagtgcacatcgccatcgccagccctgcaacat 1046  
QY 1302 ttttactgcccataatggctggacagagtggtctccgtgcacagtgacatgttgcca 1361  
Db 1047 ttttactgcccataatggctggacagagtggtctccgtgcacagtgacatgttgcca 1106  
QY 1362 gggctcagataccgtgtgtctctgtcatcgacacatcgagaatgcacacaggaggtgtg 1421  
Db 1107 gggctcagataccgtgtgtctctgtcatcgacacatcgagaatgcacacaggaggtgtg 1166  
QY 1422 tagcccaaaaacaaagcccaataaaagagaatgcatcgtaaccactccctgctataa 1481  
Db 1167 tagcccaaaaacaaagcccaataaaagagaatgcatcgtaaccactccctgctataa 1226  
QY 1482 acccaaaagaaacttccagtcagagccaagtgtccatgtgttcaaacagctcaagagct 1541  
Db 1227 acccaaaagaaacttccagtcagagccaagtgtccatgtgttcaaacagctcaagagct 1286  
QY 1542 agaagaagagagctgtgtgtcagagagccctcgttcatccagaggcctgtcggcctg 1601  
Db 1287 agaagaagagagctgtgtgtcagagagccctcgttcatccagaggcctgtcggcctg 1346  
QY 1602 cacagtcacctgtgtgtgggagaccaggtgcgaatagtcagggtccaggtgctcctgtc 1661  
Db 1347 cacagtcacctgtgtgtgggagaccaggtgcgaatagtcagggtccaggtgctcctgtc 1406

QY 1662 tttctcagtcctgtgtgtgacctgctctattgacagagtgtaagggcccaagccagcatc 1721  
Db 1407 tttctcagtcctgtgtgtgacctgctctattgacagagtgtaagggcccaagccagcatc 1466  
QY 1722 ccagcgtcctgtttatgcaggcccatgcagcggggaataattcctaagttcaaccacagaga 1781  
Db 1467 ccagcgtcctgtttatgcaggcccatgcagcggggaataattcctaagttcaaccacagaga 1526  
QY 1782 gacagatgggtcttctgtgtgacctgaggaatttcogacagctgtatgactgggagatga 1841  
Db 1527 gacagatgggtcttctgtgtgacctgaggaatttcogacagctgtatgactgggagatga 1586  
QY 1842 ggggttccaccaagtgctccagagctcgtgagaggagtgctccagaggcgtgtgtgagctg 1901  
Db 1587 ggggttccaccaagtgctccagagctcgtgagaggagtgctccagaggcgtgtgtgagctg 1646  
QY 1902 cttgaacaaacagactcgggagcctgctgagagaaacctgtgcgtgacacagcccgcc 1961  
Db 1647 cttgaacaaacagactcgggagcctgctgagagaaacctgtgcgtgacacagcccgcc 1706  
QY 1962 cccacagctcctgaagtcctgcaatttggatccctgcccagcaaggtgggaaattggcaa 2021  
Db 1707 cccacagctcctgaagtcctgcaatttggatccctgcccagcaaggtgggaaattggcaa 1766  
QY 2022 gtggagtcctatgtatgtctcacatgtggggtcgccctacagacagacagctctctctcag 2081  
Db 1767 gtggagtcctatgtatgtctcacatgtggggtcgccctacagacagacagctctctcag 1826  
QY 2082 ccaactgttctccagagagatgaatgaacagctcatcctggtgatgagctgtgtcgcga 2141  
Db 1827 ccaactgttctccagagagatgaatgaacagctcatcctggtgatgagctgtgtcgcga 1886  
QY 2142 gcccaagcccagacogtgcgaagcttgaacctttaaattgccccccagcctgtgtaccc 2201  
Db 1887 gcccaagcccagacogtgcgaagcttgaacctttaaattgccccccagcctgtgtaccc 1946  
QY 2202 tgcacagtgccagcctgttccagaaacgtgtggcgggggtgttcaagaaacgtgaggtct 2261  
Db 1947 tgcacagtgccagcctgttccagaaacgtgtggcgggggtgttcaagaaacgtgaggtct 2006  
QY 2262 ttgcaagcagcgcgtgctgctgagcttccctggagcttccctgagacattctgttcagc 2321  
Db 2007 ttgcaagcagcgcgtgctgctgagcttccctggagcttccctgagacattctgttcagc 2066  
QY 2322 ttcaaaacctgctgcagcaagcatgcgaagaagtgaactgtccagcagtggtctct 2381  
Db 2067 ttcaaaacctgctgcagcaagcatgcgaagaagtgaactgtccagcagtggtctct 2126  
QY 2382 ctccagactggacagagtggttccacaagctgcggggaaggccacccagactcgaagcgcct 2441  
Db 2127 ctccagactggacagagtggttccacaagctgcggggaaggccacccagactcgaagcgcct 2186  
QY 2442 ttgcccgaagaagtgcataaaacccgctctcaacggttgcattccacccctgtgcccgc 2501  
Db 2187 ttgcccgaagaagtgcataaaacccgctctcaacggttgcattccacccctgtgcccgc 2246  
QY 2502 cctgccttctcttctccatcagccctgtatgtgtggaacctgtgcgaagggccggcg 2561  
Db 2247 cctgccttctcttctccatcagccctgtatgtgtggaacctgtgcgaagggccggcg 2306  
QY 2562 gccatccagaagcacagccgcacatcgccgcgcaggaaggtctacatccagactcg 2621  
Db 2307 gccatccagaagcacagccgcacatcgccgcgcaggaaggtctacatccagactcg 2366  
QY 2622 caggcagaggaagctgcacttctgtgtgtggggcgtcttcgctacctgctcccaagcgc 2681  
Db 2367 caggcagaggaagctgcacttctgtgtgtggggcgtcttcgctacctgctcccaagcgc 2426  
QY 2682 ggtgtgtgtgctgctgcccggcgcgagggtccgagccctcatcaacctgggaaagga 2741  
Db 2427 ggtgtgtgtgctgctgcccggcgcgagggtccgagccctcatcaacctgggaaagga 2486

QY	2742	ggcaggaacccctatcagctgcagcagctcaagctgagcccttcggtatctcaagat	2801
Db	2487		
		cggcagcacctcatcagctgcagcagctcacgtgcccccttcggatactcaagat	2546
QY	2802	ccacgcctcaagccctcgatgcaaggcgtctacacctgctcagcgggccggcccgga	2861
Db	2547		
		ccacgcctcaagccctcgatgcaaggcgtctacacctgctcagcgggccggcccgga	2606
QY	2862	gcactttgtattaaagctcatcgagggaacccgcaagctcgtggccccgccccttagccc	2921
Db	2607		
		gcactttgtattaaagctcatcgagggaacccgcaagctcgtggccccgccccttagccc	2666
QY	2922	gagaaatgagaaagagtgcttcgcggggaaggaaggcgccgaagagagccctgcagac	2981
Db	2667		
		gagaaatgagaaagagtgcttcgcggggaaggaaggcgccgaagagagccctgcagac	2726
QY	2982	ccacaacacacagaacgagatcttccaaacggcagcaagcggaagcggggcctggc	3041
Db	2727		
		ccacaacacacagaacgagatcttccaaacggcagcaagcggaagcggggcctggc	2786
QY	3042	ggccaaacccgggagccgctaagcagacctcgtctcccgctgcttgagcagggcgctg	3101
Db	2787		
		ggccaaacccgggagccgctaagcagacctcgtctcccgctgcttgagcagggcgctg	2846
QY	3102	gcccgagagctgctggcctcgtggaggcgagactccgcggaagaaagaaacacgacctc	3161
Db	2847		
		gcccgagagctgctggcctcgtggaggcgagactctcggaaagaaacacgacctc	2906
QY	3162	ggaggagaccgggtgcagagcaagtgctctgcacctgcccctcaccatggtagccga	3221
Db	2907		
		ggaggagaccgggtgcagagcaagtgctctgcacctgcccctcaccatggtagccga	2966
QY	3222	gcagcgcgccctggagacatcctggggaacctctccagagccgagagctgcgca	3281
Db	2967		
		gcagcgcgccctggagacatcctggggaacctctccagagccgagaggtgcgca	3026
QY	3282	cctctacagcaagcacctggtggccagctggcccaggagattctccgagccacctgga	3341
Db	3027		
		cctctacagcaagcacctggtggccagctggcccagagattctcgcagccacctgga	3086
QY	3342	gcaccaggacgcctctgaagccctcgagcgaggaacttcccagtaactctctcgcc	3401
Db	3087		
		gcaccaggacgcctctgaagccctcgagcgaggaacttcccagtaactctctcgcc	3146
QY	3402	tcataaacagctgtctgcttcagcagctccccgcgacctctccacggggaagcgcg	3461
Db	3147		
		tcataaacagctgtctgcttcagcagctccccgcgacctctccacggggaagcgcg	3206
QY	3462	ggaggctctgaagccacacccgaagcccaacctctctgcgaagatctcagcgccca	3521
Db	3207		
		ggaggctctgaagccacacccgaagcccaacctctctgcgaagatctcagcgccca	3266
QY	3522	gcagctctcagcctcggaagtggtcaaccacctggggcagacggtggccctgccagcg	3581
Db	3267		
		gcagctctcagcctcggaagtggtcaaccacctggggcagacggtggccctgccagcg	3326
QY	3582	gcacctgagttcttctgcactgtgagccatcgccaccccaagcctaccalcaactg	3641
Db	3327		
		gcacctgagtg-tcttctgcactgtgagccaTeggnacccaagcctaccatcaactg	3385
QY	3642	ggccggaatgagaagaagttcaagttcaagtgacaggaattctctacagccagatgattc	3701
Db	3386		
		ggccggaatgagaagaag-tcaagttcaagtgacaggaattctctacagccagatgattc	3444
QY	3702	cttacagattctt-ggcaccagttgg-aagcagatgtgggtttctacacttgcaatgcacc	3759
Db	3445		
		yttacagattcttgggcaccagtggaagcagatgtgggtttctacacttgcaatggca-c	3503
QY	3760	aatgccttgggatacagactctgtctccattgccttcacattgaaggaagccac	3814
Db	3504		
		aatgccttgggatacagatttgcctcatttggcgtaacttgaaggaagccccc	3558

RESULT	4
AAA47458	
ID	AAA47458 standard; cDNA; 2689 BP.
XX	
AC	AAA47458;
XX	
20-OCT-2000	(first entry)
XX	
Human TANGO 224	coding sequence.
XX	
TANGO; 128;	140; 197; 212; 213; 224; 239; modulating agent; asthma;
XX	
graft versus-host diseases; rheumatoid arthritis; psoriasis;	
XX	
inflammatory bowel disease; septic shock; ulcerative colitis;	
XX	
Crohn's disease; chronic myelogenous leukemia; cancer; liver	
XX	
disease; Hodgkin's disease; osteoarthritis; Lyme's disease;	
XX	
cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;	
XX	
systemic lupus erythematosus; transgenic animal; diagnosis;	
XX	
prognosis; prophylactic; therapeutic; human; ds.	
XX	
Homo sapiens.	
XX	
Key	Location/Qualifiers
XX	
CDS	1..1443
FT	/**tag= a
FT	/product= TANGO 224
FT	
XX	
WO200039284-A1.	
XX	
06-JUL-2000.	
XX	
23-DEC-1999;	99WO-US1025.
XX	
30-DEC-1998;	98US-0223546.
XX	
(MILL-) MILLENNIUM PHARM INC.	
XX	
Holtzman DA;	
XX	
WPI: 2000-465743/40.	
XX	
P-PSDB; AAB01425.	
XX	
Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,	
XX	
224 and 239 polypeptides useful for the treatment of asthma, rheumatoid	
XX	
arthritis, psoriasis and autoimmune diseases	
XX	
Claim 1; Fig 7; 209pp; English.	
XX	
Nucleic acids encoding TANGO polypeptides are useful as modulating	
XX	
agents for regulating cellular processes like asthma, graft	
XX	
versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory	
XX	
bowel disease, septic shock, ulcerative colitis, Crohn's disease,	
XX	
chronic myelogenous leukemia, cancer, liver disease, Hodgkin's	
XX	
disease, osteoarthritis, Lyme's disease, cachexia and autoimmune	
XX	
diseases e.g. myasthenia gravis, autoimmune diabetes and systemic	
XX	
lupus erythematosus. The nucleic acids are also useful for producing	
XX	
transgenic animals and the TANGO polypeptides themselves. Partial	
XX	
TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in	
XX	
forensic biology, for diagnostic assays, prognostic assays,	
XX	
pharmacogenomics and for monitoring clinical trials. TANGO	
XX	
polypeptides are suitable for both prophylactic and therapeutic	
XX	
methods for treating a subject at risk of a disorder or having a	
XX	
disorder associated with aberrant TANGO expression. A wide range	
XX	
of cellular disorders can be treated.	
XX	
Sequence 2689 BP; 633 A; 740 C; 734 G; 582 T; 0 other;	
XX	
Query Match	48.3%; Score 2556.2; DB 21; Length 2689;
Best Local Similarity	98.9%; Pred. No. 0;
Matches 2594; Conservative	0; Mismatches 3; Indels 26; Gaps
QY	1 atggaatgcctgcgtoggcaactcctgcacactgcctcttttggctttcttgcgc 60

Db 67 atgaatgctgcgtcgggcaactccttggaacatgctctcttcttggttctctgctc 126  
QY 61 ctgagttccaggacccgacgctcggagagacccgggacgacctatggatgctggg 120  
Db 127 ctgagttccaggacccgacgctcggagagacccgggacgacctatggatgctggg 186  
QY 121 ccattgagtgaaatgctcagccacctgcgggggtggggtcctctactctctgagcgctgc 180  
Db 187 ccatggagtgaaatgctcagccacctgcggggagggggtcctctactctctgagcgctgc 246  
QY 181 ctgagcagcaagagctgtgaggaagaataatccgatacacagaacatcagtaattgtggac 240  
Db 247 ctgagcagcaagagctgtgaggaagaataatccgatacacagaacatcagtaattgtggac 306  
QY 241 tgcaccagaagcagtgatttccgagctcagcaatgctcagctcataatgatgtcaag 300  
Db 307 tgcaccagaagcagtgatttccgagctcagcaatgctcagctcataatgatgtcaag 366  
QY 301 caccatggccagtttatgaatggcttccctgtgtctaatgacccctgacacccatgttca 360  
Db 367 caccatggccagtttatgaatggcttccctgtgtctaatgacccctgacacccatgttca 426  
QY 361 ctcaagtgcacagcccaaggaacacccctggttgtgaactagcacctaaaggtcttagat 420  
Db 427 ctcaagtgcacagcccaaggaacacccctggttgtgaactagcacctaaaggtcttagat 486  
QY 421 ggtacggtgtctacagaaattttggatatgtgcacagtggtttatgccaaattgtt 480  
Db 487 ggtacggtgtctacagaaattttggatatgtgcacagtggtttatgccaaattgtt 546  
QY 481 ggtcgcgataccagctgggaacacccctgcgaagagataaactgtggggtctgcaacgga 540  
Db 547 ggtcgcgataccagctgggaacacccctgcgaagagataaactgtggggtctgcaacgga 606  
QY 541 gatgggtccacctgcggctggtccgagggcagtaataatccagctctccgcaacaaa 600  
Db 607 gatgggtccacctgcggctggtccgagggcagtaataatccagctctccgcaacaaa 666  
QY 601 tcggatgatactggttgcaattccctatggaagttagacatatcgcttcttaaaa 660  
Db 667 tcggatgatactggttgcaattccctatggaagttagacatatcgcttcttaaaa 726  
QY 661 ggtcctgatactatctggaacacacccctccaggggactaaagtgaacacagt 720  
Db 727 ggtcctgatactatctggaacacacccctccaggggactaaagtgaacacagt 786  
QY 721 ctacgctccacaggaaacttctctgtggacaattctagtggaactccagaaatttcca 780  
Db 787 ctacgctccacaggaaacttctctgtggacaattctagtggaactccagaaatttcca 846  
QY 781 gacaaagatactagaatgctggaaccactccacagcagatttccatgcaagattcgt 840  
Db 847 gacaaagatactagaatgctggaaccactccacagcagatttccatgcaagattcgt 906  
QY 841 aactcgggtcccgctgacagtcagctccagttccatctctatacaccctccacccga 900  
Db 907 aactcgggtcccgctgacagtcagctccagttccatctctatacaccctccacccga 966  
QY 901 tggaggagacggatttcttctctgctcagcaacctgtggaggaggttatcagctgaca 960  
Db 967 tggaggagacggatttcttctctgctcagcaacctgtggaggaggttatcagctgaca 1026  
QY 961 tcaggtgagtgatactgaggaacacccctggtggtgctgaccaatctgtcactat 1020  
Db 1027 tcggtgagtgatactgaggaacacccctggtggtgctgaccaatctgtcactat 1086  
QY 1021 taccagagaacatacaaccccaacccagcttcaggaagtgaacttgatcctgttcca 1080  
Db 1087 taccagagaacatacaaccccaacccagcttcaggaagtgaacttgatcctgttcca 1146  
QY 1081 gc-----cagtgacgggatacaagcagatcagcttatg 1114  
Db 1147 gccagaggttgccattattgttctaacagtgacgggatacaagcagatcagcttatg 1206

QY 1115 acctaccatcccttctcgtggtggaggccacccctatggaccgctgctctcctcgt 1174  
Db 1207 acctaccatcccttctcgtggtggaggccacccctatggaccgctgctctcctcgt 1266  
QY 1175 gtggggggggtccagagccggcgagtttctctgtgtgagagagacatccaggggcatg 1234  
Db 1267 gtggggggggtccagagccggcgagtttctctgtgtggaaggagacatccaggggcatg 1326  
QY 1235 tcaattcaagtgaagagtggaaatgcattgtacacccctaaagtgcacctcgcgcagccct 1294  
Db 1327 tcaattcaagtgaagagtggaaatgcattgtacacccctaaagtgcacctcgcgcagccct 1386  
QY 1295 gcaacatttttgactgccttaaatggctggcacaggagtgctctcgtgcacagtgcacat 1354  
Db 1387 gcaacatttttgactgccttaaatggctggcacaggagtgctctcgtgcacagtgcacat 1446  
QY 1355 gtggccagggtccagataccgtgtgtccctctgcacatccacatcgaggaatgcacacag 1414  
Db 1447 gtggccagggtccagataccgtgtgtccctctgcacatcgaggaatgcacacag 1506  
QY 1415 gaggtctagccccaaaaaagccccacataaaaggaggaatgcattcgtaccactccct 1474  
Db 1507 gaggtctagccccaaaaaagccccacataaaaggaggaatgcattcgtaccactccct 1566  
QY 1475 gctataaacccaaaagaaaacttccagtcgaggccaagtgtccatgttcaacaagctc 1534  
Db 1567 gctataaacccaaaagaaaacttccagtcgaggccaagtgtccatgttcaacaagctc 1626  
QY 1535 aagagctagaagaggagctgtgtgtgcagaggagccctcgttccatccagagccctcgt 1594  
Db 1627 aagagctagaagaggagctgtgtgtgcagaggagccctcgttccatccagagccctcgt 1686  
QY 1595 cggcctgcacagtcacactgtgtgtgtgggcccaggtgtcgaaatagtcaggtgcagagtg 1654  
Db 1687 cggcctgcacagtcacactgtgtgtgtgggcccaggtgtcgaaatagtcaggtgcagagtg 1746  
QY 1655 tctgtcttctctcagtcctcgtgtgacctgacctatgtacagagctgaaggggcccaagc 1714  
Db 1747 tctgtcttctctcagtcctcgtgtgacctgacctatgtacagagctgaaggggcccaagc 1806  
QY 1715 cagcatcccggtgctctgtatgcagggcccatgcaggggaaaattccagttcaacc 1774  
Db 1807 cagcatcccggtgctctgtatgcagggcccatgcaggggaaaattccagttcaacc 1866  
QY 1775 cagacagacagatgggtcttctgtgtggtcgcaggtattcgcagagctglatgactggg 1834  
Db 1867 cagacagacagatgggtcttctgtgtggtcgcaggtattcgcagagctglatgactggg 1926  
QY 1835 agtatgaggggttcaccaaagtcctcagagtcctgtgagagaggtgtccagagagctgtgg 1894  
Db 1927 agtatgaggggttcaccaaagtcctcagagtcctgtgagagaggtgtccagagagctgtgg 1986  
QY 1895 tgaagctgttgaacaaacagactcggagcctgctgaggaagacccctgtgcgtgaccagcc 1954  
Db 1987 tgaagctgttgaacaaacagactcggagcctgctgaggaagacccctgtgcgtgaccagcc 2046  
QY 1955 gccggcccccacagctcctgaaatcctgcaatttgatccctgccagcaagtgaggaaa 2014  
Db 2047 gccggcccccacagctcctgaaatcctgcaatttgatccctgccagcaagtgaggaaa 2106  
QY 2015 ttggcaagtggaagtcctatgtacatgtgggtcggctcagacacagagagagctct 2074  
Db 2107 ttggcaagtggaagtcctatgtacatgtgggtcggctcagacacagagagagctct 2166  
QY 2075 tctgcagccacctgcttctcagagagatgaatgaacacagtcattcctggctgatgactgt 2134  
Db 2167 tctgcagccacctgcttctcagagagatgaatgaacacagtcattcctggctgatgactgt 2226  
QY 2135 gtcgcagcccaagcccgacagcgggtgaagctgttaaccgcttgaattgccccccagct 2194  
Db 2227 gtcgcagcccaagcccgacagcgggtgaagctgttaaccgcttgaattgccccccagct 2286

Qy	2195	ggtaacctgcacagtgcagcgcgtgttccagaacatgttgccaggggtgttcagaaacgtg	2255
Db	2287	ggtaacctgcacagtgcagcgcgtgttccagaacatgttgccaggggtgttcagaaacgtg	2346
Qy	2255	aggttctttgcaagcagcagcagctgctgatgcagcttctcagagcttctcagaccttct	2314
Db	2347	aggttctttgcaagcagcagcagctgctgatgcagcttctcagagcttctcagaccttct	2406
Qy	2315	gttcagcttcaaaacctcctgcctgcagcagcatgcaagaaaagatgactgtccacgcagt	2374
Db	2407	gttcagcttcaaaacctcctgcctgcagcagcatgcaagaaaagatgactgtccacgcagt	2466
Qy	2375	ggcttctcagactggacagagtgttcocaaagctgcgggaaagcaccacgactcgaa	2434
Db	2467	ggcttctcagactggacagagtgttcocaaagctgcgggaaagcaccacgactcgaa	2526
Qy	2435	gcgcatttgcgaaagatgctgaaacccggcctctcaacggttgcattccaccctgt	2494
Db	2527	gcgcatttgcgaaagatgctgaaacccggcctctcaacggttgcattccaccctgt	2586
Qy	2495	gccgcacctgcctttcttctcctccatcaggccctgtatgctggcaacctgtgcaagc	2554
Db	2587	gccgcacctgcctttcttctcctccatcaggccctgtatgctggcaacctgtgcaagc	2646
Qy	2555	ccggcgcccatccacgaagcacagcccgacatcgcgccgc	2597
Db	2647	ccggcgcccatccacgaagcacagcccgacatcgcgccgc	2689
RESULT <sup>†</sup> 5			
AAA47482			
ID	AAA47482 standard; cDNA; 2689 BP.		
AC	AAA47482;		
XX			
XX			
DT	20-OCT-2000 (first entry)		
DE	Human TANGO 224 coding sequence (form 1).		
XX			
KW	TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;		
KW	graft versus-host diseases; rheumatoid arthritis; psoriasis;		
KW	inflammatory bowel disease; septic shock; ulcerative colitis;		
KW	Crohn's disease; chronic myelogenous leukemia; cancer; liver		
KW	disease; Hodgkin's disease; osteoarthritis; Lyme's disease;		
KW	cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;		
KW	systemic lupus erythematosus; transgenic animal; diagnosis;		
XX	prognosis; prophylactic; therapeutic; human; ds.		
OS	Homo sapiens.		
XX			
FH	Key		
CDS	Location/Qualifiers		
FT	67..2688		
FT	/*tag= a		
FT	/product= TANGO 224		
XX			
PN	WO200039284-A1.		
XX			
PD	06-JUL-2000.		
XX			
PF	23-DEC-1999; 99WO-US31025.		
XX			
PR	30-DEC-1998; 98US-0223546.		
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
XX			
PI	Holtzman DA;		
XX			
DR	WPI; 2000-465743/40.		
DR	P-PSDB; AAB01431.		
XX			
PT	Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,		
PT	224 and 239 polypeptides useful for the treatment of asthma, rheumatoid		
PT	arthritis, psoriasis and autoimmune diseases		

XX	Claim 1; Fig 30; 209pp; English.	
PS		
XX		
CC	Nucleic acids encoding TANGO polypeptides are useful as modulating	
CC	agents for regulating cellular processes like asthma, graft	
CC	versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory	
CC	bowel disease, septic shock, ulcerative colitis, Crohn's disease,	
CC	chronic myelogenous leukemia, cancer, liver disease, Hodgkin's	
CC	disease, osteoarthritis, Lyme's disease, cachexia and autoimmune	
CC	diseases e.g. myasthenia gravis, autoimmune diabetes and systemic	
CC	lupus erythematosus. The nucleic acids are also useful for producing	
CC	transgenic animals and the TANGO polypeptides themselves. Partial	
CC	TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in	
CC	forensic biology, for diagnostic assays, prognostic assays,	
CC	pharmacogenomics and for monitoring clinical trials. TANGO	
CC	polypeptides are suitable for both prophylactic and therapeutic	
CC	methods for treating a subject at risk of a disorder or having a	
CC	disorder associated with aberrant TANGO expression. A wide range	
CC	of cellular disorders can be treated.	
XX		
SQ	Sequence 2689 BP; 633 A; 739 C; 734 G; 583 T; 0 other;	
	Query Match 48.3%; Score 2552.6; DB 21; Length 2689;	
	Best Local Similarity 98.8%; Pred. No. 0;	
	Matches 2592; Conservative 0; Mismatches 4; Indels 27; Gaps	
QY	1 atggaatgctccctcaggcaactctcggcacactgctctcttcttggtttccgtgc 60	
DB	67 atggaatgctccctcaggcaactctcggcacactgctctcttcttggtttccgtgc 126	
QY	61 ctgagttccaggaaccgacgctccgagagagaccggagcgcctatggatgctggggc 120	
DB	127 ctgagttccaggaaccgacgctccgagagagaccggagcgcctatggatgctggggc 186	
QY	121 ccatggagtgaaatgctcacgcacctcgcgggggtgggcctctactctctgagcgctgc 180	
DB	187 ccatggagtgaaatgctcacgcacctcgcgggggtgggcctctactctgagcgctgc 246	
QY	181 ctgagcagcaagagctgtgaaggagaataatccgatacagaacatgcagtaatgtggac 240	
DB	247 ctgagcagcaagagctgtgaaggagaataatccgatacagaacatgcagtaatgtggac 306	
QY	241 tgcaccacgaagcaggtgattccgagctcagcaatgctcagctcataatgatgtcaag 300	
DB	307 tgcaccacgaagcaggtgattccgagctcagcaatgctcagctcataatgatgtcaag 366	
QY	301 caccatggccagttttatgaatggcttctctgtgtctaatacgacctgacaccccatgttca 360	
DB	367 caccatggccagttttatgaatggcttctctgtgtctaatacgacctgacaccccatgttca 426	
QY	361 ctcaagtgtccaaagccaaaggaacacacctggttgtgaactagcaccctaaaggtcttagat 420	
DB	427 ctcaagtgtccaaagccaaaggaacacacctggttgtgaactagcaccctaaaggtcttagat 486	
QY	421 ggtacgcggtgtctatacagaatctttggatatgtgcacgtggtgttatccaaattgtt 480	
DB	487 ggtacgcggtgtctatacagaatctttggatatgtgcacgtggtgttatccaaattgtt 546	
QY	481 ggcctgcgatcacaccagctgggaagcacccgtcgaaggagaataaactgtggggtctgcaacgga 540	
DB	547 ggcctgcgatcacaccagctgggaagcacccgtcgaaggagaataaactgtggggtctgcaacgga 606	
QY	541 gatgggtccacctgcacgctggtccgagggcagtgataataatccacagctctccgacaacaa 600	
DB	607 gatgggtccacctgcacgctggtccgagggcagtgataataatccacagctctccgacaacaa 666	
QY	601 tcgcatgatactggtgtgcaattccctatggaaagttagacatatatgccttcttcttaaa 660	
DB	667 tcgcatgatactggtgtgcaattccctatggaaagttagacatatatgccttcttcttaaa 726	
QY	661 ggtcctgatcacttatctctggaaaccaaaccctccaggggactaaagtgtgaaacagt 720	



KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;  
 KW muscular dystrophy; ss.  
 XX Homo sapiens.  
 OS  
 PN W0200000610-A2.  
 XX  
 PD 06-JAN-2000.  
 XX  
 XX 25-JUN-1999; 99WO-US14484.  
 XX  
 XX 26-JUN-1998; 98US-0090762.  
 PR 31-JUL-1998; 98US-0094983.  
 PR 01-OCT-1998; 98US-0102686.  
 PR 11-DEC-1998; 98US-0112129.  
 XX  
 XX (INCY-) INCYTE PHARM INC.  
 XX  
 XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KU, Baughn MR;  
 PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;  
 PI Bandman O;  
 XX  
 DR WPI; 2000-160673/14.  
 DR P-PSDB; AAY87347.  
 XX  
 XX New human signal peptide-containing proteins useful in treatment,  
 PT prevention and diagnosis of e.g. cancer, inflammation and  
 PT cardiovascular disease -  
 PT  
 XX Claim 9; Page 321-322; 327pp; English.  
 XX  
 CC AAY98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the  
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have  
 CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,  
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can  
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders  
 CC associated with decreased activity or function of HSP. Antagonists of  
 CC HSP are used to treat or prevent disorders associated with increased  
 CC activity or function of HSP. Such disorders include cell proliferation  
 CC (including cancer), inflammation, cardiovascular, neurological,  
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,  
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,  
 CC asthma, Crohn's disease, microbial or other infections, congestive or  
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP  
 CC nucleic acids can be used for the recombinant production of HSP, for  
 CC detecting HSP in standard hybridisation and amplification assays (for  
 CC diagnosis and monitoring), in gene therapy, as antisense,  
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences  
 CC or genetic variations, and for chromosomal mapping. HSP are also used to  
 CC raise specific antibodies (Ab) and to screen for agonists and  
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic  
 CC antagonists, in competitive drug screens, and for purification of HSP  
 CC from natural sources.  
 XX  
 SQ Sequence 3045 BP; 763 A; 772 C; 768 G; 723 T; 19 other;

QY	121	ccatggagtgatgctcaccacacccgcggggtgggctcctactctctgagcgctgc	180
DB	250	ccatggagtgatgctcaccacacccgcggggtgggctcctactctctgagcgctgc	309
QY	181	ctgagcagcagagctgtgaaggagaataatccgatacacagaacatgagtaattgtgac	240
DB	310	ctgagcagcagagctgtgaaggagaataatccgatacacagaacatgagtaattgtgac	369
QY	241	tgccaccagcagcaggtgattccgagctcagcaatgctcagctcataatgattgacg	300
DB	370	tgccaccagcagcaggtgattccgagctcagcaatgctcagctcataatgattgacg	429
QY	301	caccatggccagttttatgaatggctctgtctaatgacctgacacccatgttcca	360
DB	430	caccatggccagttttatgaatggctctgtctaatgacctgacacccatgttcca	489
QY	361	ctcaagtgcacagccaaagggaacacccctggtgtgaactagcacctaaagctcttagat	420
DB	490	ctcaagtgcacagccaaagggaacacccctggtgtgaactagcacctaaagctcttagat	549
QY	421	ggtacgcgttgctatcacagaatctttggatgtgcatcagtggtgttatgccaattgtt	480
DB	550	ggtacgcgttgctatcacagaatctttggatgtgcatcagtggtgttatgccaattgtt	609
QY	481	ggctgcgataccagctgggaagcaccgtccaaggaagataaactgtgggtctcgcacgga	540
DB	610	ggctgcgataccagctgggaagcaccgtccaaggaagataaactgtgggtctcgcacgga	669
QY	541	gatgggtccacactgccgctggtccgagggcagatataaactccagctctccgacacaaa	600
DB	670	gatgggtccacactgccgctggtccgagggcagatataaactccagctctccgacacaaa	729
QY	601	tcggatgatactgtggtgtaattccctatggaagttagcatatttcgcttcttaaaa	660
DB	730	tcggatgatactgtggtgtaattccctatggaagttagcatatttcgcttcttaaaa	789
QY	661	ggtcctgatactatataatgaaacccacccctccagggagactaaaggtaaaacagt	720
DB	790	ggtcctgatactatataatgaaacccacccctccagggagactaaaggtaaaacagt	849
QY	721	ctcagctccacaggaacttctctgtggacaattctagtgtgactccagaaattcca	780
DB	850	ctcagctccacaggaacttctctgtggacaattctagtgtgactccagaaattcca	909
QY	781	gacaaagatactgagaatggctggaccaactcagcagagatttcattgcaagattcgt	840
DB	910	gacaaagatactgagaatggctggaccaactcagcagagatttcattgcaagattcgt	969
QY	841	aactcgggtccgctcagatcagatccagttctctctatcaacccatcaccacga	900
DB	970	aactcgggtccgctcagatcagatccagttctctctatcaacccatcaccacga	1029
QY	901	tgagggagacgagatttcttctcctcagcaacctgtggaggaggttatcagctgaca	960
DB	1030	tgagggagacgagatttcttctcctcagcaacctgtggaggaggttatcagctgaca	1089
QY	961	tcggctgagtgctacgatactgaggaacacccgtgtgtgtgctgacaaactactcactat	1020
DB	1090	tcggctgagtgctacgatactgaggaacacccgtgtgtgtgctgacaaactactcactat	1149
QY	1021	taccagagacatcaaaccccaaacccagcttcagagtgcaacttggaactcttgcca	1080
DB	1150	taccagagacatcaaaccccaaacccagcttcagagtgcaacttggaactcttgcca	1209
QY	1081	gccagtgcaggtatacagcagatcattccttatgacctaccatccctctcctcggtg	1140
DB	1210	gccagtgcaggtatacagcagatcattccttatgacctaccatccctctcctcggtg	1269
QY	1141	gagggccacccatgacccggtgctcctcctcgtgtgtgtgggggggcatccagagccggca	1200
DB	1270	gagggccacccatgacccggtgctcctcctcgtgtgtgtgggggggcatccagagccggca	1329
QY	1201	gtttcctgtgtgtgagggagacatccaggggcatgtcattcagtggaagatggaaatgc	1260

```
Db 1330 gtctctgtgtgagagagacatccagggagcatgtcaacttcagtggagagtggaatgc 1389
|||||
QY 1261 atgtacacccttaagatgcccacgcgcagccctgcacacatttttgaactgcccctaaatgg 1320
|||||
Db 1390 atgtacacccttaagatgcccacgcgcagccctgcacacatttttgaactgcccctaaatgg 1449
|||||
QY 1321 ctggacacagagtggtctccgtgcacagtgacaatgtggccaggccctcagataccctgtg 1380
|||||
Db 1450 ctggacacagagtggtctccgtgcacagtgacaatgtggccaggccctcagataccctgtg 1509
|||||
QY 1381 gtccctgtcatgcacatcgaggaatgcacacagaggtgtgagcccaaaaaacaaagccc 1440
|||||
Db 1510 gtccctgtcatgcacatcgaggaatgcacacagaggtgtgagcccaaaaaacaaagccc 1569
|||||
QY 1441 cacataaagaggaatgcacatcgaccactccctgctataaaccacaaagagaaactcca 1500
|||||
Db 1570 cacataaagaggaatgcacatcgaccactccctgctataaaccacaaagagaaactcca 1629
|||||
QY 1501 gtccagagccaagtcccatggttcaaaacaaagctcaagagctagaagaagagctgctgtg 1560
|||||
Db 1630 gtccagagccaagtcccatggttcaaaacaaagctcaagagctagaagaagagctgctgtg 1689
|||||
QY 1561 tcagaggagccctcgttcaaccagagccctggtgcgctgcacagtgcaactgtggtgtg 1620
|||||
Db 1690 tcagaggagccctcgttcaaccagagccctggtgcgctgcacagtgcaactgtggtgtg 1749
|||||
QY 1621 gggacccaggtgcgaatagtcaggtgcaggtgctccgtgctcttctcagtcctggtgct 1680
|||||
Db 1750 gggacccaggtgcgaatagtcaggtgcaggtgctccgtgctcttctcagtcctggtgct 1809
|||||
QY 1681 gacctgctattgacagagtggaaggcccaagccagcagcaccagctcgtgctgttatgca 1740
|||||
Db 1810 gacctgctattgacagagtggaaggcccaagccagcagcaccagctcgtgctgttatgca 1869
|||||
QY 1741 gggccatcagcggggaaattccctgagttcaaccagacagacagatgggctctttggt 1800
|||||
Db 1870 gggccatcagcggggaaattccctgagttcaaccagacagacagatgggctctttggt 1929
|||||
QY 1801 gggctgaagatttcagcagagctgatgactggagtgagtgagggttaccacaaagtctcc 1860
|||||
Db 1930 gggctgaagatttcagcagagctgatgactggagtgagtgagggttaccacaaagtctcc 1989
|||||
QY 1861 gagtctgtgagcaggtgtccagagagctgtgtgagctccttgaaacaaacagactcgg 1920
|||||
Db 1990 gagtctgtgagcaggtgtccagagagctgtgtgagctccttgaaacaaacagactcgg 2049
|||||
QY 1921 gagcc-tgctgagagaaacctgtgctgacacagccgcggcccccacagctcctgaagtc 1979
|||||
Db 2050 gagcctgtcgtgaggaacctgtgctgagcca-cgcgcggccccacagctcctgaagtc 2108
|||||
QY 1980 ctgcaatttgatccctgcccagaag 2006
|||||
Db 2109 ctgcaatttgatccctgcccagaag 2135
|||||

RESULT 7
ABAO6676
ID ABAO6676 standard; cDNA; 2031 BP.
XX
AC ABAO6676;
XX
DT 10-JAN-2002 (first entry)
DE Human cDNA SEQ ID NO: 342.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
```

PR	29-SEP-2000;	2000US-236321P
PR	29-SEP-2000;	2000US-236367P
PR	29-SEP-2000;	2000US-236369P
PR	29-SEP-2000;	2000US-236369P
PR	29-SEP-2000;	2000US-236370P
PR	29-SEP-2000;	2000US-236370P
PR	02-OCT-2000;	2000US-236802P
PR	02-OCT-2000;	2000US-237037P
PR	02-OCT-2000;	2000US-237038P
PR	02-OCT-2000;	2000US-237039P
PR	02-OCT-2000;	2000US-237040P
PR	13-OCT-2000;	2000US-239935P
PR	13-OCT-2000;	2000US-239937P
PR	20-OCT-2000;	2000US-240960P
PR	20-OCT-2000;	2000US-241221P
PR	20-OCT-2000;	2000US-241785P
PR	20-OCT-2000;	2000US-241786P
PR	20-OCT-2000;	2000US-241787P
PR	20-OCT-2000;	2000US-241808P
PR	20-OCT-2000;	2000US-241809P
PR	20-OCT-2000;	2000US-241826P
PR	01-NOV-2000;	2000US-244517P
PR	08-NOV-2000;	2000US-246475P
PR	08-NOV-2000;	2000US-246475P
PR	08-NOV-2000;	2000US-246476P
PR	08-NOV-2000;	2000US-246477P
PR	08-NOV-2000;	2000US-246478P
PR	08-NOV-2000;	2000US-246523P
PR	08-NOV-2000;	2000US-246524P
PR	08-NOV-2000;	2000US-246525P
PR	08-NOV-2000;	2000US-246526P
PR	08-NOV-2000;	2000US-246527P
PR	08-NOV-2000;	2000US-246528P
PR	08-NOV-2000;	2000US-246532P
PR	08-NOV-2000;	2000US-246609P
PR	08-NOV-2000;	2000US-246610P
PR	08-NOV-2000;	2000US-246611P
PR	08-NOV-2000;	2000US-246613P
PR	17-NOV-2000;	2000US-249207P
PR	17-NOV-2000;	2000US-249208P
PR	17-NOV-2000;	2000US-249209P
PR	17-NOV-2000;	2000US-249210P
PR	17-NOV-2000;	2000US-249211P
PR	17-NOV-2000;	2000US-249212P
PR	17-NOV-2000;	2000US-249213P
PR	17-NOV-2000;	2000US-249214P
PR	17-NOV-2000;	2000US-249215P
PR	17-NOV-2000;	2000US-249216P
PR	17-NOV-2000;	2000US-249217P
PR	17-NOV-2000;	2000US-249218P
PR	17-NOV-2000;	2000US-249244P
PR	17-NOV-2000;	2000US-249245P
PR	17-NOV-2000;	2000US-249264P
PR	17-NOV-2000;	2000US-249265P
PR	17-NOV-2000;	2000US-249297P
PR	17-NOV-2000;	2000US-249299P
PR	17-NOV-2000;	2000US-249300P
PR	01-DEC-2000;	2000US-250160P
PR	01-DEC-2000;	2000US-250391P
PR	05-DEC-2000;	2000US-251030P
PR	05-DEC-2000;	2000US-251988P
PR	05-DEC-2000;	2000US-256719P
PR	06-DEC-2000;	2000US-251479P
PR	08-DEC-2000;	2000US-251856P
PR	08-DEC-2000;	2000US-251868P
PR	08-DEC-2000;	2000US-251869P
PR	08-DEC-2000;	2000US-251989P
PR	11-DEC-2000;	2000US-251990P
PR	11-DEC-2000;	2000US-254097P
PR	05-JAN-2001;	2001US-259678P
XX	(HUMA-)	HUMAN GENOME SCI INST
PA	Rosen CA,	Barash SC, Ruben
XX		
XX		

XX	WPI: 2001-476161/51.
DR	P-PSDB: ABB10454.
XX	
PT	Isolated nucleic acid molecule encoding an inflammation-associated polypeptide is used in preventing, treating or ameliorating a medical condition -
PT	
XX	
XX	Claim 1; SEQ ID NO: 342; 859pp + Sequence Listing; English.
PS	
XX	The present invention provides human cDNAs, proteins and related genomic DNAs. These can be used in the treatment of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders and inflammation. The present sequence is a cDNA of the invention.
XX	
SQ	Sequence 2031 BP; 449 A; 610 C; 564 G; 408 T; 0 other;
	Query Match 36.2%; Score 1912.8; DB 22; Length 2031;
	Best Local Similarity 99.4%; Pred. No. 0;
	Matches 1920; Conservative 0; Mismatches 12; Indels 0; Gaps
QY	3358 ctgaagccctcggagcgaggactcccagtgactcttcgctcataaacacgtgctt 3417 
Dd	9 ctgcaggaaatcgccagcaggactcccagtgactcttcgctcataaacacgtgctt 68 
QY	3418 ggttcacgacgtccctcggacctctccacggggagccggggaggtctcgaag 3477 
Dd	69 ggcttcacgacgtccctcggacctctccacggggagccggggaggtctcgaag 128 
QY	3478 ccacaccgcccaagccccaccatctcgccaagtatccagggccacagctctcagctcg 3537 
Dd	129 ccacaccgcccaagccccaccatctcgccaagtatccagggccacagctctcagctcg 188 
QY	3538 gaagtgtcacccactgggcacacgtagtgccctgccaggggacactgagtgcttt 3597 
Dd	189 gaggtgtgtcacccactgggcacacgtagtgccctgccaggggacactgagtgcttt 248 
QY	3598 ctgcactgtgagcccatggccaccaaggctaccatcagctgggcaggaaatggagaa 3657 
Dd	249 ctgcactgtgagcccatggccaccaaggctaccatcagctgggcaggaaatggagaa 308 
QY	3658 gaagttcagttcagtgacaggaattcttacagccagatgatctctacagatcttgcaa 3717 
Dd	309 gaagttcagttcagtgacaggaattcttacagccagatgatctctacagatcttgcaa 368 
QY	3718 ccagtggaagcagatgtgggtttcacactgtcaatgccacaatgctctgggatacagc 3777 
Dd	369 ccagtggaagcagatgtgggtttcacactgtcaatgccacaatgctctgggatacagc 428 
QY	3778 ttgttgtcattggcgtcacattacaggaagccactagtgaaaaacgtcacgaatgaca 3837 
Dd	429 ttgttgtcattggcgtcacattacaggaagccactagtgaaaaacgtcacgaatgaca 488 
QY	3838 gtgatcaacacaggagaagcctcagctcacgtataggaagcacacatcaaaaacgtg 3897 
Dd	489 gtgatcaacacaggagaagcctcagctcacgtataggaagcacacatcaaaaacgtg 548 
QY	3898 cagggagtgaaatgtgacaatacaactgccaggttgccagagtgccctgaagctgaagtcaact 3957 
Dd	549 cagggagtgaaatgtgacaatacaactgccaggttgccagagtgccctgaagctgaagtcaact 608 
QY	3958 tgggttcaggaataaaagcaaaactggctccccgccaccatctgcacgaaggtcctctgctg 4017 
Dd	609 tgggttcaggaataaaagcaaaactggctccccgccaccatctgcacgaaggtcctctgctg 668 
QY	4018 ctcaaaaacgtgtccctctcgatcagggccctgtactcctcagggcgcccaattcttcact 4077 
Dd	669 ctcaaaaacgtgtccctctcgatcagggccctgtactcctcagggcgcccaattcttcact 728 
QY	4078 ggagagctgactgagagaccacagctgctgatctctagatccccccacagtcaccacacag 4137 



Db 729 ggagagctgaactgagagcaccagctgctgatctctagatcccccccaagccccacacag 788  
QY 4138 ttggaagacatcaggcccttgcctgctgcactggaccgaacctctctcaagtgcagc 4197  
Db 789 ttggaagacatcaggcccttgcctgctgcactggaccgaacctctctcaagtgcagc 848  
QY 4198 tctctctgggaacacagctgctcctggatcctcctggaaattctgctcctctgctgctccc 4257  
Db 849 tctctctgggaacacagctgctcctggatcctcctggaaattctgctcctctgctgctccc 908  
QY 4258 atcaaaaggtccacctgctccatataccttggtttcatgggtgagcgaacattgtcaact 4317  
Db 909 atcaaaaggtccacctgctccatataccttggtttcatgggtgagcgaacattgtcaact 968  
QY 4318 gccacaggactgacgcacatccttgccagctggacagatccttcaagtgcacacactt 4377  
Db 969 gccacaggactgacgcacatccttgccagctggacagatccttcaagtgcacacactt 1028  
QY 4378 agcgtggttctcaagggaattcagctccttgcctgtcagaatgagcagggtgtctcatg 4437  
Db 1029 agcgtggttctcaagggaattcagctccttgcctgtcagaatgagcagggtgtctcatg 1088  
QY 4438 cagaaggcatctttagtatcaaatgactggtgtctgtgacagactgcaaacctgc 4497  
Db 1089 cagaaggcatctttagtatcaaatgactggtgtctgtgacagactgcaaacctgc 1148  
QY 4498 tcagcctcctgtgttaacgggggttcagcagcccccttgagtgctcgtgacacgc 4557  
Db 1149 tcagcctcctgtgttaacgggggttcagcagcccccttgagtgctcgtgacacgc 1208  
QY 4558 acggaggttcaacctgccactgcagcagggaaggttcgctcctgcgtgacccccatgcg 4617  
Db 1209 acggaggttcaacctgccactgcagcagggaaggttcgctcctgcgtgacccccatgcg 1268  
QY 4618 tgcacccggagagactgccccttctcgttggatggtgacctcctgctgctgtaccccg 4677  
Db 1269 tgcacccggagagactgccccttctcgttggatggtgacctcctgctgctgtaccccg 1328  
QY 4678 agctgtggggagaggtgtccagaccgcgggtgacctgtcaaaagctgaaagcctctggg 4737  
Db 1329 agctgtggggagaggtgtccagaccgcgggtgacctgtcaaaagctgaaagcctctggg 1388  
QY 4738 atctccacctctgttccaatgacatgtgcacccaggtgcacccagcctgtggacacc 4797  
Db 1389 atctccacctctgttccaatgacatgtgcacccaggtgcacccagcctgtggacacc 1448  
QY 4798 caggcctgtaacacagcagctgtgtgtggagtgggccttctccagctggggccagtgaat 4857  
Db 1449 caggcctgtaacacagcagctgtgtgtggagtgggccttctccagctggggccagtgaat 1508  
QY 4858 gggccttgcatcgggcctcaacctagctgtgcaacacagacaaagtcttctgccagacag 4917  
Db 1509 gggccttgcatcgggcctcaacctagctgtgcaacacagacaaagtcttctgccagacag 1568  
QY 4918 gatgcatccttaccatcagacagcagctgagtgctcttcagcagcctgtgagcaccag 4977  
Db 1569 gatgcatccttaccatcagacagcagctgagtgctcttcagcagcctgtgagcaccag 1628  
QY 4978 aactgtgttgtagaggcctgcagtacactggagagtcagcctgtgacccctgtgcaca 5037  
Db 1629 aactgtgttgtagaggcctgcagtacactggagagtcagcctgtgacccctgtgcaca 1688  
QY 5038 gctacctgtggcaactacgcttccagtcctccgctgtggagtggtgcatgcccgcacc 5097  
Db 1689 gctacctgtggcaactacgcttccagtcctccgctgtggagtggtgcatgcccgcacc 1748  
QY 5098 acaaggcagctgactgagcactgtctcctggggcccccgcctgccactggcagcgc 5157  
Db 1749 acaaggcagctgactgagcactgtctcctggggcccccgcctgccactggcagcgc 1808  
QY 5158 tgcacatcaccccatgtgaaacatggagtgacagacacccaccaggtactgcgagaag 5217  
Db 1809 tgcacatcaccccatgtgaaacatggagtgacagacacccaccaggtactgcgagaag 1868

QY 5218 gtcaaacagctgaactctgccaaactcagccagtttaaatctcgtcgtggaacttgt 5277  
Db 1869 gtcaaacagctgaactctgccaaactcagccagtttaaatctcgtcgtggaacttgt 1928  
QY 5278 ggcaaaagcgtga 5289  
Db 1929 ggcaaaagcgtga 1940  
RESULT 8  
AAZ34146  
ID AAZ34146 standard; cDNA; 1869 bp.  
XX  
AC AAZ34146;  
XX  
DT 07-DEC-1999 (first entry)  
XX Human PRO1071 nucleotide sequence.  
XX  
DE  
XX  
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein; ss.  
XX Homo sapiens.  
PN WO9946281-A2.  
XX  
PD 16-SEP-1999.  
XX  
PF 08-MAR-1999; 99WO-US05028.  
XX  
PR 10-MAR-1998; 98US-0077450.  
PR 11-MAR-1998; 98US-0077632.  
PR 11-MAR-1998; 98US-0077641.  
PR 11-MAR-1998; 98US-0077649.  
PR 12-MAR-1998; 98US-0077791.  
PR 13-MAR-1998; 98US-0078004.  
PR 17-MAR-1998; 98US-0040220.  
PR 20-MAR-1998; 98US-0078886.  
PR 20-MAR-1998; 98US-0078910.  
PR 20-MAR-1998; 98US-0078936.  
PR 20-MAR-1998; 98US-0078939.  
PR 25-MAR-1998; 98US-0079294.  
PR 26-MAR-1998; 98US-0079656.  
PR 27-MAR-1998; 98US-0079663.  
PR 27-MAR-1998; 98US-0079664.  
PR 27-MAR-1998; 98US-0079689.  
PR 27-MAR-1998; 98US-0079728.  
PR 27-MAR-1998; 98US-0079786.  
PR 30-MAR-1998; 98US-0079920.  
PR 30-MAR-1998; 98US-0079923.  
PR 31-MAR-1998; 98US-0080105.  
PR 31-MAR-1998; 98US-0080107.  
PR 31-MAR-1998; 98US-0080165.  
PR 31-MAR-1998; 98US-0080194.  
PR 01-APR-1998; 98US-0080327.  
PR 01-APR-1998; 98US-0080328.  
PR 01-APR-1998; 98US-0080333.  
PR 01-APR-1998; 98US-0080334.  
PR 08-APR-1998; 98US-0081049.  
PR 08-APR-1998; 98US-0081070.  
PR 08-APR-1998; 98US-0081071.  
PR 09-APR-1998; 98US-0081195.  
PR 09-APR-1998; 98US-0081203.  
PR 09-APR-1998; 98US-0081229.  
PR 15-APR-1998; 98US-0081817.  
PR 15-APR-1998; 98US-0081838.  
PR 15-APR-1998; 98US-0081952.  
PR 15-APR-1998; 98US-0081955.  
PR 21-APR-1998; 98US-0082568.  
PR 21-APR-1998; 98US-0082569.  
PR 22-APR-1998; 98US-0082700.



```

Db 1093 tcggctgagtgctacgatctgaggagcaaccgtgtgtgtgtgctgacccaatactgtcaactat 1152
QY 1021 taccagagaacatacaaaacccaacccaagcttcaggagtgcaacttggatccttgcga 1080
Db 1153 taccagagaacatacaaaacccaacccaagcttcaggagtgcaacttggatccttgcga 1212
QY 1081 gccagtgacggatatacaagcagatcagcttctatgacctacacatcccttctcgtgtg 1140
Db 1213 gccagtgacggatatacaagcagatcagcttctatgacctacacatcccttctcgtgtg 1272
QY 1141 gaggccacccccatggaccgctgtcctcctcgtgtgggggggacccagagccgggca 1200
Db 1273 gaggccacccccatggaccgctgtcctcctcgtgtgggggggacccagagccgggca 1332
QY 1201 gtttcctgtgtgagagagacatccaggggcagtgcaactcagtggaagtggaatgc 1260
Db 1333 gtttcctgtgtgagagagacatccaggggcagtgcaactcagtggaagtggaatgc 1392
QY 1261 atgtcacacccctaaagatgccatcgcgcagccctgcacacattttgactgacctaaatgg 1320
Db 1393 atgtcacacccctaaagatgccatcgcgcagccctgcacacattttgactgacctaaatgg 1452
QY 1321 ctggcacaggadgtgtcctcgtgcagtgacatgtggccagggctcagatcacogtgg 1380
Db 1453 ctggcacaggadgtgtcctcgtgcagtgacatgtggccagggctcagatcacogtgg 1512
QY 1381 gtcctctgcatcgacatcaggaatgacacagagagcgtgtagcccaaaaacaaagccc 1440
Db 1513 gtcctctgcatcgacatcaggaatgacacagagagcgtgtagcccaaaaacaaagccc 1572
QY 1441 cacataaagagggaatgcacgtaccctcctcctcgtataaacccaagagaaattcca 1500
Db 1573 cacataaagagggaatgcacgtaccctcctcctcgtataaacccaagagaaattcca 1632
QY 1501 gtcaggcccaagtggtccatggttcaaaacagctcaagagctagaagaaggagctgtg 1560
Db 1633 gtcaggcccaagtggtccatggttcaaaacagctcaagagctagaagaaggagctgtg 1692
QY 1561 tcagagagagccctgctatccagagagcctggtgg 1597
Db 1693 tcagagagagccctgctatggttgaagacagactg 1729

RESULT 9
AAC78532
ID AAC78532 standard; cDNA; 1869 BP.
XX
AC AAC78532;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PR01071 (UNQ528) nucleotide sequence SEQ ID NO:300.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW expressed sequence tag; detection; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.

```

```

PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan MJ;
PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI; 2000-611443/58.
DR
DR P-PSDB; AAB44285.
XX
PT Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
XX Claim 2; Fig 119; 636pp; English.
XX
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
XX sequence tag) sequences which encode secreted or transmembrane PRO
XX polypeptides. The PRO polynucleotides and polypeptides have cytostatic
XX activity. The polynucleotides and polypeptides can be used for detecting
XX the presence of PRO polypeptides in samples, for linking bioactive
XX molecules to cells and for modulating biological activities of cells,
XX using the polypeptides for specific targeting. The polypeptide targeting
XX can be used to kill the target cells, e.g. for the treatment of cancers.
XX The polypeptide pairs provide specific targeting of bioactive molecules
XX to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
XX the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

```

```

Query Match 29.7%; Score 1573; DB 21; Length 1869;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1582; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 atggaatgctgcgtcgggcaactcctgacactgctcctcttcttctggttctctgctc 60
Db 133 atggaatgctgcgtcgggcaactcctgacactgctcctcttcttctggttctctgctc 192

QY 61 ctgagttccaggacgcacgtccgagagagagagagagagagagagagagagagagagagag 120
Db 193 ctgagttccaggacgcacgtccgagagagagagagagagagagagagagagagagagagag 252

QY 121 ccattggagtgaatgctcagcacctgcgggggtggggtggggtggggtggggtggggtggg 180
Db 253 ccattggagtgaatgctcagcacctgcgggggtggggtggggtggggtggggtggggtggg 312

QY 181 ctgagagcagagagctgtgaaggagaataatccgatacagaacatcagtaattgtgac 240
Db 313 ctgagcagcagagctgtgaaggagaataatccgatacagaacatcagtaattgtgac 372

QY 241 tgccaccagaacaggtgatttcagagctcagcaatgctcagctcaataatgatgtcaag 300
Db 373 tgccaccagaacaggtgatttcagagctcagcaatgctcagctcaataatgatgtcaag 432

QY 301 caccatggccagtttgaatgaatgcttctctgtcttaataatgacctgacacccaatgtca 360
Db 433 caccatggccagtttgaatgaatgcttctctgtcttaataatgacctgacacccaatgtca 492

QY 361 ctcaatgccaagccaaggaacaaacccctggttgaactagcacctagacacaggtcttagat 420
Db 493 ctcaatgccaagccaaggaacaaacccctggttgaactagcacctagacacaggtcttagat 552

```

```
Qy 421 ggtacgctgtctatcacagaatttttgatatgtgcatcagtggtttatgccaattgtt 480
Db 553 ggtacgctgtctatcacagaatttttgatatgtgcatcagtggtttatgccaattgtt 612
Qy 481 ggtcgtcatcaccagctgggaagcaccgtcagaagaataaactgtggggtctgcaacgga 540
Db 613 ggtcgtcatcaccagctgggaagcaccgtcagaagaataaactgtggggtctgcaacgga 672
Qy 541 gatgggtccacctgccggtcgtgctcgaaggcaggtatataatccagctctcgcgaacaaa 600
Db 673 gatgggtccacctgccggtcgtgctcgaaggcaggtatataatccagctctcgcgaacaaa 732
Qy 601 tcgatatgatactgtggttgcaattccctatggaagtagacatatctgctgtcttaaaa 660
Db 733 tcggatgatactgtggttgcaattccctatggaagtagacatatctgctgtcttaaaa 792
Qy 661 ggtctgatcacttatctgtaaaccaaaaacccctccaggggactaaagggtgaaacagt 720
Db 793 ggtctgatcacttatctgtaaaccaaaaacccctccaggggactaaagggtgaaacagt 852
Qy 721 ctcagctccacaggaactttcctgtgtgacaattctagtggaacttcagaaatttcca 780
Db 853 ctcagctccacaggaactttcctgtgtgacaattctagtggaacttcagaaatttcca 912
Qy 781 gacaaagagatactgagatggtgtgacacactcacagcagatttcatgttcaagattcgt 840
Db 913 gacaaagagatactgagatggtgtgacacactcacagcagatttcatgttcaagattcgt 972
Qy 841 aactcgggtccgctgacagtagacagtagacagtagacagtagacagtagacagtagacag 900
Db 973 aactcgggtccgctgacagtagacagtagacagtagacagtagacagtagacagtagacag 1032
Qy 901 tggaggagagacgattttctctgtctcagcaacctgtggaggaggttatcagctgaca 960
Db 1033 tggaggagagacgattttctctgtctcagcaacctgtggaggaggttatcagctgaca 1092
Qy 961 tcggctgagtgctacagatcgtgaggagcaaccgtgtggttgcctgacccaatactgtcaactat 1020
Db 1093 tcggctgagtgctacagatcgtgaggagcaaccgtgtggttgcctgacccaatactgtcaactat 1152
Qy 1021 taccagagaacatacaaaacccaacccaagcttcagagtagtgaacttgatcctgttcca 1080
Db 1153 taccagagaacatacaaaacccaacccaagcttcagagtagtgaacttgatcctgttcca 1212
Qy 1081 gccagtgcagtagacagcagatcgtcttatgacctacacatcccttccctcctcgttg 1140
Db 1213 gccagtgcagtagacagcagatcgtcttatgacctacacatcccttccctcctcgttg 1272
Qy 1141 gaggccaccccatgaccgctgctcctcctcctcgtgtggtgggggggcatccagagccgggca 1200
Db 1273 gaggccaccccatgaccgctgctcctcctcctcgtgtggtgggggggcatccagagccgggca 1332
Qy 1201 gtttctctgtgaggagagacatccaggggcatgctcaacttcagtggaagagtgaagtgc 1260
Db 1333 gtttctctgtgaggagagacatccaggggcatgctcaacttcagtggaagagtgaagtgc 1392
Qy 1261 atgtacacccctaaagtgcacatcgccagccctgcacacatcttttgaactgccttaaatgg 1320
Db 1393 atgtacacccctaaagtgcacatcgccagccctgcacacatcttttgaactgccttaaatgg 1452
Qy 1321 ctggcagcagagtggtctcgtgacagtgacatgtggccaggggcctcagataccgtgtg 1380
Db 1453 ctggcagcagagtggtctcgtgacagtgacatgtggccaggggcctcagataccgtgtg 1512
Qy 1381 gtctctgtacacacatcgaggaatgacacagggagcgtgtgaccccaaaaacaaagccc 1440
Db 1513 gtctctgtacacacatcgaggaatgacacagggagcgtgtgaccccaaaaacaaagccc 1572
Qy 1441 cacataaagaggaatgcatctgaccactccctgctataaccccaagagaacttcca 1500
Db 1573 cacataaagaggaatgcatctgaccactccctgctataaccccaagagaacttcca 1632
```

Qy 1501 gtcagaggccaagtgtccatgttccaaacaaagctcaagagctagaagaaggagctctgtg 1560  
Db 1633 gtcagaggccaagtgtccatgttccaaacaaagctcaagagctagaagaaggagctctgtg 1692

Qy 1561 tcagaggagcctcttccatccagagagcctggtcgg 1597  
Db 1693 tcagaggagcctctgtaagtgtgtaaaagcacagactg 1729

RESULT 10  
AAS46142  
ID AAS46142 standard; cDNA; 1869 BP.  
XX  
AC AAS46142;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human DNA encoding PRO polypeptide sequence #218.  
XX  
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;  
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;  
KW PCR primer.  
XX  
OS Homo sapiens.  
XX  
PN WO200168848-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 28-FEB-2001; 2001WO-US06520.  
XX  
PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 03-MAR-2000; 2000US-187202P.  
PR 06-MAR-2000; 2000US-186968P.  
PR 14-MAR-2000; 2000US-189320P.  
PR 14-MAR-2000; 2000US-189328P.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 21-MAR-2000; 2000US-190828P.  
PR 21-MAR-2000; 2000US-191007P.  
PR 21-MAR-2000; 2000US-191048P.  
PR 21-MAR-2000; 2000US-191314P.  
PR 28-MAR-2000; 2000US-192655P.  
PR 29-MAR-2000; 2000US-193032P.  
PR 29-MAR-2000; 2000US-193053P.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 04-APR-2000; 2000US-194449P.  
PR 04-APR-2000; 2000US-194647P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-196000P.  
PR 11-APR-2000; 2000US-196187P.  
PR 11-APR-2000; 2000US-196690P.  
PR 11-APR-2000; 2000US-196820P.  
PR 18-APR-2000; 2000US-198121P.  
PR 18-APR-2000; 2000US-198585P.  
PR 25-APR-2000; 2000US-199397P.  
PR 25-APR-2000; 2000US-199550P.  
PR 25-APR-2000; 2000US-199654P.  
PR 03-MAY-2000; 2000US-201516P.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 05-JUN-2000; 2000US-209832P.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 22-AUG-2000; 2000US-064488.  
PR 22-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
XX



## RESULT 11

AAFG3447

ID AAF63447 standard; cDNA; 1803 BP.

XX AC

XX AC

XX AC

DT 14-MAY-2001 (first entry)

XX XX

DE Human ADAMTS-R1 cDNA sequence.

XX XX

ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;  
 tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;  
 Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;  
 metastasis; embryogenesis; egg implantation; ADAMTS-RI; ss.

XX OS

XX Homo sapiens.

XX W0200111074-A2.

PN 15-FEB-2001.

PD 03-AUG-2000; 2000WO-US21223.

PF 06-AUG-1999; 99US-0369364.

PR (CLEV-) CLEVELAND CLINIC FOUND.

XX (APTE/) APTE S S.

PA (HURS/) HURSKAINEN T L.

PA (HIRO/) HIROHATA S.

XX XX

XX Apte SS, Hurskainen TL, Hirohata S;

PI WPI; 2001-159978/16.

XX P-PSDB; AAB72290.

DR Murine and human 'A Disintegrin-like And Metalloprotease domain with

Thrombospondin type I motifs' proteins and the nucleic acids encoding

them, useful for treating e.g. tumours, inflammation and arthritis -

XX Claim 15; Fig 11; 181pp; English.

XX XX

This invention relates to murine and human ADAMTS-N (A disintegrin-like  
 and metalloprotease domain with thrombospondin type I motifs) proteins,  
 designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the  
 invention are cDNA sequences encoding the proteins, and antibodies  
 specific for the proteins. The nucleic acid sequences and proteins may be  
 used in the prevention, diagnosis and treatment of diseases associated  
 with inappropriate ADAMTS-N expression. Disorders that may be treated  
 using the nucleic acids, proteins and antibodies include, for example  
 tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos  
 syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage  
 in arthritic (both inflammatory and non-inflammatory) disease,  
 angiogenesis, tumour growth and metastases, and they may also be used for  
 controlling embryogenesis and implantation of fertilised eggs. The  
 present sequence represents cDNA encoding human ADAMTS-RI.

XX Sequence 1803 BP; 475 A; 460 C; 457 G; 411 T; 0 other;

XX Query Match 29.7%; Score 1571.4; DB 22; Length 1803;

XX Best Local Similarity 99.0%; Pred. No. 0;

XX Matches 1581; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 atgaatgctgcggtcggaactcctggcacactgtctctcttctggtctctgctc 60

Db 51 atgaatgctgcggtcggaactcctggcacactgtctctcttctggtctctgctc 110

QY 61 ctgagttccagagaccgacgtccgagagagaccgagacggccctatggatgcctggggc 120

Db 111 ctgagttccagagaccgacgtccgagagagaccgagacggccctatggatgcctggggc 170

QY 121 ccatggagtgatgctacgcacctgcgggggtggggtcctctctctctgagggcgtgc 180

Db 171 ccatggagtgatgctacgcacctgcgggggtggggcgccaaactctctgagggcgtgc 230

QY 181 ctgagcagaagagctgtgaaggaagaataatccgatacagacaacatgcaagtaattgagac 240

Db 231 ctgagcagaagagctgtgaaggaagaataatccgatacagacaacatgcaagtaattgagac 290

QY 241 tggccaccagaagcaggtgatttccgagagtcaggaatgctcagtcataatgctcaag 300

Db 291 tggccaccagaagcaggtgatttccgagagtcaggaatgctcagtcataatgctcaag 350

QY 301 caccatggccatttatgaatggttctctgtcttaataagccctgacacaccatgttca 360

Db 351 caccatggccatttatgaatggttctctgtcttaataagccctgacacaccatgttca 410

QY 361 ctcaagtgccaaagcgaagaaacacccctggttggtaactagcaccctaaggtcttagat 420

Db 411 ctcaagtgccaaagcgaagaaacacccctggttggtaactagcaccctaaggtcttagat 470

QY 421 ggtacgcgtgtctatacagaatctttggatatgtgcacgtggtttatgccaaattgtt 480

Db 471 ggtacgcgtgtctatacagaatctttggatatgtgcacgtggtttatgccaaattgtt 530

QY 481 ggtcgcgtacacagctgggaagcacgcgtcaaggaagataaactgtgggtctgcacacgga 540

Db 531 ggtcgcgtacacagctgggaagcacgcgtcaaggaagataaactgtgggtctgcacacgga 590

QY 541 gatgggtccacctgcgggtggttccgagggcagataaaatcccagctctccgcaacaaa 600

Db 591 gatgggtccacctgcgggtggttccgagggcagataaaatcccagctctccgcaacaaa 650

QY 601 tcggatgatactgtgttgcgaatccctatggaagtagacacattcgtctcttaaaa 660

Db 651 tcggatgatactgtgttgcgaatccctatggaagtagacacattcgtctcttaaaa 710

QY 661 ggtcctgatactatctggaacacaaacccctcaggggactaaagtgaacacagt 720

Db 711 ggtcctgatactatctggaacacaaacccctcaggggactaaagtgaacacagt 770

QY 721 ctacgtctccacagaaacttctctgttggaacaattctagtgtggactccagaaattcca 780

Db 771 ctacgtctccacagaaacttctctgttggaacaattctagtgtggactccagaaattcca 830

QY 781 gacaaagatactagaaatggctggaccactcagacagatttcattgtccaagattcgt 840

Db 831 gacaaagatactagaaatggctggaccactcagacagatttcattgtccaagattcgt 890

QY 841 aactcgggtccgcgtgacagtcacagtcacagttcattcttatacaccatcatccacga 900

Db 891 aactcgggtccgcgtgacagtcacagtcacagttcattcttatacaccatcatccacga 950

QY 901 tggaggagagcggatttcttctgtctcagcaacctgtgagagaggttatcagctgaca 960

Db 951 tggaggagagcggatttcttctgtctcagcaacctgtgagagaggttatcagctgaca 1010

QY 961 tcgctgagtgctacagatctgagagcaacgtgtggtgctgacccaactgtcactat 1020

Db 1011 tcgctgagtgctacagatctgagagcaacgtgtggtgctgacccaactgtcactat 1070

QY 1021 taccagagaacatcaaaccccaaaccccaagcttcaggaggtgcaactgtgattctgtcca 1080

Db 1071 taccagagaacatcaaaccccaaaccccaagcttcaggaggtgcaactgtgattctgtcca 1130

QY 1081 gccagtgacgatacaaacagacatgactgactatgactctaccatccctctcctcggtgg 1140

Db 1131 gccagtgacgatacaaacagacatgactgactatgactctaccatccctctcctcggtgg 1190

QY 1141 gaggccaccccatgaccgcgtgctcctcctggtggggggggaatccagagcgggca 1200

Db 1191 gaggccaccccatgaccgcgtgctcctcctggtggggggggaatccagagcgggca 1250

QY 1201 gttctcgtgtgtggagagagacatccagggggaatgctcaacttcagtggagagtggaatgc 1260

Db 1251 gttctctgtgtggaggaggacatccaggggcatgtoacttcagtgaagagtggaatgc 1310  
QY 1261 atgtacacccctaagatgccatgcgcagccctgcagccatttttgactgcccataatgg 1320  
Db 1311 atgtacacccctaagatgccatgcgcagccctgcagccatttttgactgcccataatgg 1370  
QY 1321 ctggcacagagtggtctcctcggtcacagtgacatgtggtgcccaggccctcagataccgtgtg 1380  
Db 1371 ctggcacagagtggtctcctcggtcacagtgacatgtggtgcccaggccctcagataccgtgtg 1430  
QY 1381 gtctctgcatcaccatcgagggaatgcacacagagaggtgtgtagcccaaaaacaaagccc 1440  
Db 1431 gtctctgcatcaccatcgagggaatgcacacagagaggtgtgtagcccaaaaacaaagccc 1490  
QY 1441 cacataaaagagaatgcacatcgtaaccactccctgctatataaaacccaaagagaacttcca 1500  
Db 1491 cacataaaagagaatgcacatcgtaaccactccctgctatataaaacccaaagagaacttcca 1550  
QY 1501 gtcgagggcgaagtgtccatggttcaacaagctcaagagctagaagagagctgtgtg 1560  
Db 1551 gtcgagggcgaagtgtccatggttcaacaagctcaagagagctagaagagagctgtgtg 1610  
QY 1561 tcagagagccctgttcattccaccagagggtgtgctgg 1597  
Db 1611 tcagagagccctgttaagtgttaaaagcacagactg 1647

## RESULT 12

AAI66429

ID AAI66429 standard; cDNA; 1616 BP.

XX AC AAI66429;

XX DT 04-DEC-2001 (first entry)

XX DE Human thrombotic protein 46 coding sequence.

XX KW Human; thrombotic protein 46; cancer; HIV infection; gene therapy; ss.

XX OS Homo sapiens.

XX PN CN1300774-A.

XX PD 27-JUN-2001.

XX PF 22-DEC-1999; 99CN-0125685.

XX PR 22-DEC-1999; 99CN-0125685.

XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX PI Mao Y, Xie Y;

XX DR WPT; 2001-530462/59.

XX DR P-PSDB; AAG78679.

XX PT New human thrombotic protein 46 and polynucleotide encoding it, useful for treating cancer and human immunodeficiency virus infection -

XX PS Claim 6; Page 24-25(Disclosure); 32pp; Chinese.

XX CC The present invention provides the protein and coding sequences of human thrombotic protein 46. The sequences can be used in the treatment of cancer and HIV infection. The present sequence is the coding sequence of the invention.

XX SQ Sequence 1616 BP; 386 A; 438 C; 426 G; 366 T; 0 other;

## Query Match

Best Local Similarity 25.9%; Score 1369.2; DB 22; Length 1616;

Matches 1374; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 638 gacatattcgcttctttaaaggctcctgatacttatctatctggaaccccaaacctcc 697  
Db 1 gacatattcgcttctttaaaggctcctgatacttatctatctggaaccccaaacctcc 60  
QY 698 agggactaaagggtgaaacagctctcagctccacaggaacttctctgtgacaaattcta 757  
Db 61 agggactaaagggtgaaacagctcagctccacaggaacttctctgtgacaaattcta 120  
QY 758 gtgtggacttccagaataattccagacaagagatactgagaaatggtgagacactcacag 817  
Db 121 gtgtggacttccagaataattccagacaagagatactgagaaatggtgagacactcacag 180  
QY 818 cagatttcattgtcaagattcgttaactcgggtccgctgacagtgacagtgacagtgacatct 877  
Db 181 cagatttcattgtcaagattcgttaactcgggtccgctgacagtgacagtgacagtgacatct 240  
QY 878 tctatcaaccatcatccaccgatggaggagacggattttcttctctgtccagaaacct 937  
Db 241 tctatcaaccatcatccaccgatggaggagacggattttcttctctgtccagaaacct 300  
QY 938 gtggaggaggttatcagctgacatcggctgagtgctacgatctgagagcaacogtgtgg 997  
Db 301 gtggaggaggttatcagctgacatcggctgagtgctacgatctgagagcaacogtgtgg 360  
QY 998 ttgtgaccaatactgtcactattaccagagaacatcaaaaccccaaccccaagcttcagg 1057  
Db 361 ttgtgaccaatactgtcactattaccagagaacatcaaaaccccaaccccaagcttcagg 420  
QY 1058 agtgcaacttggatccttctccagccagtgacggatacaagcagatcatccttatgacc 1117  
Db 421 agtgcaacttggatccttctccagccagtgacggatacaagcagatcatccttatgacc 480  
QY 1118 tctaccatcccttctcgtggaggagcccccattggaccgctgctcctcctcgtgtg 1177  
Db 481 tctaccatcccttctcgtggaggagcccccattggaccgctgctcctcctcgtgtg 540  
QY 1178 gggggggcatccagagccggcagtttctgtgtggagagacatccaggggcagtgca 1237  
Db 541 gggggggcatccagagccggcagtttctgtgtggagagacatccaggggcagtgca 600  
QY 1238 ctccagtgaagagtggaatgcattacacccctaaagtgcacccatgagcctcgcagccctgca 1297  
Db 601 ctccagtgaagagtggaatgcattacacccctaaagtgcacccatgagcctcgcagccctgca 660  
QY 1298 acattttgacitgccctaaatggtggcagagtggtctcgtgcacagtgacatgtg 1357  
Db 661 acattttgacitgccctaaatggtggcagagtggtctcgtgcacagtgacatgtg 720  
QY 1358 gccagggctcagataccgtgtggtcctctgcatcgaccatcgaggaatgcacacaggag 1417  
Db 721 gccagggctcagataccgtgtggtcctctgcatcgaccatcgaggaatgcacacaggag 780  
QY 1418 gctgtagcccaaaaacaaagcccccataaaagaggaatgcattacccactccctgct 1477  
Db 781 gctgtagcccaaaaacaaagcccccataaaagaggaatgcattacccactccctgct 840  
QY 1478 ataaacccaaagagaacttccagtcgagggccaaagtgtgcattggttcaaacagctcaag 1537  
Db 841 ataaacccaaagagaacttccagtcgagggccaaagtgtgcattggttcaaacagctcaag 900  
QY 1538 agctagaagaagagctgctgtgtcagagagccctcgttccatccagaggcctggtcgg 1597  
Db 901 agctagaagaagagctgctgtgtcagagagccctcgttccatccagaggcctggtcgg 960  
QY 1598 cctgcacagtcacctgtgtgtggggaccagtcgcaatagtcaggtgcagagtgctcc 1657  
Db 961 cctgcacagtcacctgtgtgtggggaccagtcgcaatagtcaggtgcagagtgctcc 1020  
QY 1658 tgtctttctcagtcctgctgacctgacatgtgacagagtgtaaaagggcccaagccag 1717  
Db 1021 tgtctttctcagtcctgctgacctgacatgtgacagagtgtaaaagggcccaagccag 1080











QY 2100 gatgaataaacaagctacactcctggctgagctgtgtgcccagcccaagcccagcacggt 2159  
 Db 2333 -----cagggagaccctgcccctcctgagagtgccgagatgaaagcccacatgcttt 2387  
 QY 2160 gcaagcttgaaccgtttaattgccccccagcctggtaccctgcacagtgagcagcgtg 2219  
 Db 2388 acaagcatgcaatcagtttgaactcctcctggtggtgacattggaagatggcagcagtg 2447  
 QY 2220 ttccagaacgtgtgcccggggtgttcagaaacgtgaggttcttggcaagcagcgcagtgc 2279  
 Db 2448 ttccaggactgtggcggggggaactcagacaagagagtcacactgtcggcagctgctaac 2507  
 QY 2280 tgatggcagcttccctggagcttctcagacacttctgttccagcttcaaaaacctgctgcca 2339  
 Db 2508 ggatggcagcttttgaatctcagatgaattgagcaaggacccaaggcatcgtctca 2567  
 QY 2340 gcaagcatcacaagaagatgactgtcccagcaggtggttctcagactgacagagtg 2399  
 Db 2568 caagtctgtgccaggacagactgtcctccacatttagctgtgggagactggtcgaagtg 2627  
 QY 2400 ttccacaagctgcggggaagccaccagactcgaagcgccatttgcgaaagatgctgaa 2459  
 Db 2628 ttctgcagttgtgtgtgtggaatccagagaagaagcaggtgtgtcaaggctggcagc 2687  
 QY 2460 aacccgctctcaacggttgcgaattccacccctgtgcccgccttgccttctctctc 2519  
 Db 2688 caaaggtcggcgcatccccctcagtgagatgagtgagggatctaccagggttccctct 2747  
 QY 2520 catcagggccctgtatctggc-----aacctgtgcaaggcccggcgccatc 2567  
 Db 2748 tgtaagatcttgcagatgctgagtgagtgcaagtaaaatcaaatcagagatgaagacaaact 2807  
 QY 2568 caggaagcacagcccacatcgcgcgcggaagaggtctacatccagactcgcaggca 2627  
 Db 2808 tggtagcaggggtcccgagatcctcagtgatccagagagtgatcattcagacaagggaaga 2867  
 QY 2628 gaggaagctgcaactcgtgtgtgggggttgcctcactcgtctcccacagacggcgtggt 2687  
 Db 2868 gaagcgtattaaactgacctgacctggtgagcagcctatttgcgcccaacacatccgtgat 2927  
 QY 2688 gctgcgtgcccggcgaggttcgcgaagccctcactcactgggagagacgcca 2747  
 Db 2928 tattaagtgcctcgtgcagattccagaaatctctgatccagtgaggagagatggccg 2987  
 QY 2748 gcacctcatcagctgcagcacgttcacggtggcccttgcgctatctcaagatccaccg 2807  
 Db 2988 ttgcctgcagaactccaaacggcttggcatcaccagtcaggctcactaaaaatccacgg 3047  
 QY 2808 cctcaagccctcggatgcagggctgtcactcactgctcagcgggcccggggagacatt 2867  
 Db 3048 tcttgcctcccgcagacatcggtgtaccggtgcatgtcaggtctgtcacaggaaacagt 3107  
 QY 2868 tgtgattaaagctcatcgaggcaaccgcaagctcgtgcccggcc 2912  
 Db 3108 tgtgtcaagctcattggtactgacaaccggctcactcgcagcggcc 3152

Search completed: July 24, 2002, 02:24:47  
 Job time: 11016 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 24, 2002, 04:22:27 ; Search time 25.27 Seconds  
(without alignments)  
2699.796 Million cell updates/sec

Title: US-10-044-807-2

Perfect score: 9588

Sequence: 1 MFCRRATPGTLLFLALL.....LKLQLSQPKSRCCGTCGKA 1762

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	972	10.1	1629	1	Q9p2n4 homo sapien
2	894	9.3	1593	1	P58397 homo sapien
3	807	8.4	1077	1	Q9h324 homo sapien
4	651	6.8	1201	1	O15072 homo sapien
5	639.5	6.7	1205	1	P79331 b adamts-2
6	613	6.4	967	1	Q9wug1 rattus norv
7	608.5	6.3	1211	1	O95450 h adamts-2
8	597	6.2	968	1	P77857 mus musculu
9	584.5	6.1	967	1	O9uh18 homo sapien
10	564.5	5.9	450	1	P58459 mus musculu
11	506	5.3	860	1	Q9ukp5 homo sapien
12	495.5	5.2	890	1	Q9up79 homo sapien
13	474.5	4.9	905	1	P57110 mus musculu
14	469.5	4.9	930	1	Q9r001 mus musculu
15	462.5	4.8	930	1	O9una0 homo sapien
16	422	4.4	837	1	O75173 homo sapien
17	407.5	4.3	997	1	Q9esp7 homo sapien
18	401	4.2	630	1	P98160 homo sapien
19	376	3.9	4393	1	O15746 homo sapien
20	332.5	3.5	1906	1	Q05695 rattus norv
21	328	3.4	1914	1	O14514 homo sapien
22	326	3.4	3707	1	O60241 homo sapien
23	320.5	3.3	1259	1	P27918 homo sapien
24	317	3.3	1584	1	P43146 homo sapien
25	312.5	3.3	1572	1	P11627 mus musculu
26	311.5	3.2	469	1	P32004 homo sapien
27	311.5	3.2	1447	1	P35331 gallus gall
28	307.5	3.2	1260	1	P70211 mus musculu
29	299	3.1	1257	1	P11680 mus musculu
30	299	3.1	1284	1	P20241 drosophila
31	297.5	3.1	1447	1	
32	290.5	3.0	437	1	
33	289.5	3.0	1302	1	

34 282 2.9 1077 1 SM5A\_MOUSE Q62217 mus musculu  
35 279 2.9 1074 1 SM5A\_HUMAN Q13591 homo sapien  
36 278.5 2.9 5376 1 ZAN\_MOUSE O88799 mus musculu  
37 274.5 2.8 1040 1 AXOL\_HUMAN Q02246 homo sapien  
38 267 2.8 2012 1 DSCA\_HUMAN O60469 homo sapien  
39 265 2.8 1051 1 PTK7\_CHICK Q91048 gallus gall  
40 264 2.8 807 1 FSPQ\_RAT P33446 rattus norv  
41 258.5 2.7 1040 1 AXOL\_RAT P22063 rattus norv  
42 257 2.7 470 1 PROP\_CAVPO Q64181 cavia porce  
43 253.5 2.6 1093 1 SM5B\_MOUSE O60519 mus musculu  
44 252.5 2.6 1170 1 TSP1\_BOVIN Q28178 bos taurus  
45 250.5 2.6 1897 1 PTPF\_HUMAN P10586 homo sapien

## ALIGNMENTS

RESULT 1  
ID AT9\_HUMAN STANDARD; PRT; 1629 AA.  
AC Q9P2N4; Q9NR29;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).  
GN ADAMTS9 OR KIAA1312.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RC TISSUE=Fetal;  
RX MEDLINE=203961138; PubMed=10936055;  
RA Clark M.E.; Kellner G.S.; Turbeville L.A.; Boyer A.; Arden K.A.,  
RA Maki R.A.;  
RT \*ADAMTS 9, a novel member of the ADAM-TS/Metallospondin gene  
RT family\*;  
RL Genomics 67:343-350(2000).  
[2]  
SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).  
RP TISSUE=Brain;  
RC MEDLINE=20181126; PubMed=10718198;  
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;  
RT \*Prediction of the coding sequences of unidentified human genes. XVI.  
RT The complete sequences of 150 new cDNA clones from brain which code  
RL for large proteins in vitro.\*;  
RL DNA Res. 7:65-73(2000).  
CC -! COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -! SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
CC -! TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.  
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,  
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN  
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR  
CC THYMUS.  
CC -! DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY  
CC SIMILARITY).  
CC -! PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -! SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -! SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC - the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

Query Match 10.1%; Score 972; DB 1; Length 1629;  
Best Local Similarity 25.4%; Pred. No. 4.4e-52;  
Matches 306; Conservative 152; Mismatches 391; Indels 354; Gaps 55;

[illegible]

QY	200	KSDDTVVAIPYSGRHRLVL-----KGPDHLYLETKLOQTKGNSLSSTGTFUV-----	249
Db	764	YGNTVRIAGATNIDVRQHSFSGETDDNYL--ALSSSKGFLLN--GNFYVIMAKR	818
QY	250	----DNSSVDQKPPDK-EILRMAGPLTADFIKIRNSGSA-DSTVOFIFQPIIHR----	300
Db	819	EIRIGNAVVEYSGSETAVERINSTDRIOELLLOVLGVSKLYNDVRYSFNIPIEDKPOQ	878
QY	301	--WRETDFFPCSATCGGYQLTSAECYDLRNRV-----VADQYCHYIPENIKPK	348
Db	879	FYWNS-----HGPWQACSPKQGERKKLVCTRESQLTVSDQCRQLPO---PG	925
QY	349	PKLOECNLID-----PCPASD--GYKOIMPYDLY-----	374
Db	926	HITPCGTDCLRWIHVSRSCSAQCGIYRTL---DIYCAKYSRLDGKTEKVDGFCSS	982
QY	375	HPLPR-----WEATPWITACSSCGGIGIOBRAVSCV--EEDIQGHVTSVEBWK	419
Db	983	HPKPSNRKCKSGECNMGWRYSAMTECSKSCDGGTORRAICVNIRDV-----	1036
QY	420	CWYTPKMPIAQPCNITFDCPKLAEWSPCTVTCGGGLRYRVVLC-IDIRGHMHTGCSPKT	478
Db	1037	CTHQEKVIT-QRCSEFPQKSGDWSECLVTCGKGHKHRQWVCQFEDRLNDRMCDPET	1095
QY	479	KPHIKKECIVTPCYKPEKLVPKQAELEBGAASERPSFPIEAWSACTVC	538
Db	1096	KPTSMQTCQP-----ECASWQAGPWGQCSVTC	1123
QY	539	GVGTQVIRVRCOVLLSFSQSAD-----LPID--ECEGPK-----PASORACYAG	581
Db	1124	GQGYQIRAVKC--IIGTYSVDDNDCNAATPTQDCELPFSCHPPPAAPETRASTYSA	1181
QY	582	POSGEIPENPDETGLFGGLQDFDELYDWEYFTGFKSESCGGGVQEAUVVSLNKQTR	641
Db	1182	PRT-----QWREGSWTPCSATCKGKTRMYVSC-RDENGSG	1215
QY	642	PAEENLCVTSRRPPQLKSCNLDPCPARWEICKWSPCSILTCGVGLQTRDVFV----	697
Db	1216	VADESACATLPR-PVAKECSVTPC-GQWKALDWSSCSVTCGGGRATQVMCVNYSDHVI	1273
QY	698	SR-EMNETVILADELCRQPKPSTVQACNRPNCPP-----	730
Db	1274	DRSECDQDI-----PETDQDCSMSPCQRTPDGSLAQHPQONEDYRPSRASP	1323
QY	731	-----ANYPAWQPCSRTCGGGVQKREVLCKORMADGSLFELPFTFCSAS-----	777
Db	1324	THVLGNQWRTGPGWACSTCAGSGORRVVCOO-----ENGYTANDCVERIKP	1372
QY	778	ACQOAKDDDCPSEWLLSDWTEGSTCGEBGTOFTRSAICRMLKATGLSTVVNVTLCPLP	837
Db	1373	DEQRACESGPCP-QWAYGNWGECKLCGGGIRTLVVCQRS-----NGERFFDLSC	1422
QY	838	SSSIRPCMILATCARPGRPSTKHSPHTAAARKVYIQTQRQKLHVVVGFPAYLLPKTAVL	897
Db	1423	EILDKPPDRBQC-----NTHACPHDAA-----WSTG-----PWSSCSV	1455
QY	898	RCPARVRKPLITWEKDQGHLLSS--THVTVPAGYGLKIH--RLKPSDAGVY-----	948
Db	1456	SCRGHKQRNVYCMKADGSHLESYCKHL-AKPHGRKCRGCRPKWKAGAWSCSVSG	1514
QY	949	AGPAREHFVITLIGGNPKLV-----ARPLS-PRSEEVLACRKGKPKALQIOT-----	994
Db	1515	RGVOQRHVGCQI--GTHKIARETECNPTTRPESRDC-----QGPCRPLYTWRAEWQEC	1567
QY	995	-----HKHONGTFSNGAEKRLGAANPGSRVYDDLVSRLLE-----QGGWPGFELLA	1040
Db	1568	TKYCGESRIRKVVCDVDNKNVHGARCDSKRPVDRESCSLQPCYVWIIGEW--SEVP	1625
QY	1041	SWE	1043
Db	1626	SWE	1628

RESULT 2  
AT12\_HUMAN STANDARD; PRT; 1593 AA.  
AC P58397;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-12 precursor (BC 3.4.24.-) (A disintegrin and  
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-  
TS12).  
DE metallopeptidase.  
GN ADAMTS12.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RS SEQUENCE FROM N.A.  
RC TISSUE=Fetal lung;  
RX MEDLINE=21264577; PubMed=11279086;  
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;  
RA 'Identification, characterization, and intracellular processing of  
RT ADAM-TS12, a novel human disintegrin with a complex structural  
RT organization involving multiple thrombospondin-1 repeats.';  
RL J. Biol. Chem. 276:17932-17940(2001).  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely  
CC expressed in gastric carcinomas and in cancer cells of diverse  
CC origin.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY  
CC SIMILARITY).  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE.  
CC -!- PTM: IS SUBJECT TO AN INTRACELLULAR MATURATION PROCESS LEADING  
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE  
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND  
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1  
CC DOMAINS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ250725; CAC20419.1; -.  
CC MIM; 606184; -.  
DR PROSITE; PS50215; ADAM\_MEPRO; 1.  
DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
DR PROSITE; PS50092; TSP1; 2.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix.  
FT SIGNAL 1 25 POTENTIAL.  
FT PROPEP 26 240 BY SIMILARITY.  
FT CHAIN 241 1593 ADAMTS-12.  
FT DOMAIN 465 544 DISINTEGRIN-LIKE.  
FT DOMAIN 545 596 TSP TYPE-1 1.  
FT DOMAIN 597 700 CYS-RICH.  
FT DOMAIN 701 826 SPACER 1.  
FT DOMAIN 827 881 TSP TYPE-1 2.  
FT DOMAIN 886 943 TSP TYPE-1 3.  
FT DOMAIN 947 995 TSP TYPE-1 4.  
FT DOMAIN 996 1315 SPACER 2.  
FT DOMAIN 1316 1364 TSP TYPE-1 5.  
FT DOMAIN 1367 1423 TSP TYPE-1 6.

DOMAIN 1426 1471  
DOMAIN 1426 1471  
SITE 302 305  
METAL 208 208  
ACT\_SITE 392 392  
METAL 393 393  
METAL 396 396  
METAL 402 402  
CARBOHYD 105 105  
CARBOHYD 125 125  
CARBOHYD 215 215  
CARBOHYD 485 485  
CARBOHYD 685 685  
CARBOHYD 790 790  
CARBOHYD 951 951  
CARBOHYD 1104 1104  
CARBOHYD 1275 1275  
CARBOHYD 1300 1300  
CARBOHYD 1320 1320  
CARBOHYD 1371 1371  
CARBOHYD 1378 1378  
CARBOHYD 1503 1503  
SEQUENCE 1593 AA; 177545 MW; 07E9F48E63BD83A3 CRC64;  
  
Query Match 9.3%; Score 894; DB 1; Length 1593;  
Best Local Similarity 20.0%; Pred. No. 2.8e-47;  
Matches 358; Conservative 169; Mismatches 451; Indels 812; Gaps 67;  
  
QY 27 RSEDRDGLMDAMPWSECSTCGGASYSLRR-----LSSKSCGSRNRYRTCSNVD 80  
Db 536 KKPSIFGGWRSPWSHCSRTCGAGVQSAERLCNNPKEGKCYCTGKRYRLCNVHP 595  
QY 81 CPPAGCFRAQCSAHNDVKHGGFYEWLPVSNDDNPSLCKCOAKGTTLVVLPKVLVD 140  
Db 596 CRSEAPTFRQOCSEFTVPYKELYHWFIFN-PAHPCELYCRPIDQFSEKMLDAVID 654  
QY 141 GTRCYT--ESLDMCISGLCOLVGDHOLGSTVKEGDCGVCNGDSTCLRVGQKKSLSA 198  
Db 655 GTPCFEGSNRNVCINGICAKWGCYDEISNATEDRGVCLGDSGSCQTVRKMKFKEGS 714  
QY 199 TKSDDTVAIPYGRSHIRLVKGPDLHLYLETKTLQGTGKENSLSSTGTELVDSNVDFQK 258  
Db 715 GYVD--IGLPKKGARDIRV-----EIEGAGNFLAIRS-----E 746  
QY 259 FPDKEILRMAGPLTADFIKIRNSGSDSVQTFYQPIIHRWRETDFFPCSATCGGGY 318  
Db 747 DPEKYYL-----NGG-----FIIQW-----NGYK 766  
QY 319 LTSACQYDLRSNRVADQYCHYYPENIKPKPKLQECNLDPCPASDGYKQIMPYDLHPLP 378  
Db 767 LAGT-----VFQYDRKGDLE 781  
QY 379 RWEATPWATSSCGGGIQSRAYSCVEEDTQGHVTSVEENKCMYTPKMPAQAQCNIFDCP 438  
Db 782 KLMAI-----GPTNESV----- 793  
QY 439 KWLQAQWSPCTVTTCGGGLRYRVVLCIDHRMHVGGSPKTKPHIKBECIVPTFCYKPKK 498  
Db 794 -WIQ-----LLFQV-----INPGIKYEYTI----- 812  
QY 499 LPVEAKLPWFKQAQLEEGAANVSEEPSFIPENASACTVTCGVTQVIRVRCQVLLSFSQS 558  
Db 813 ----- 812  
QY 559 VADLPIDCEGPKPASQACYSAGEIPEFNPDETDLGLFGSLQDFDELYDWEYEGFTK 618  
Db 813 -----QKGLD---NDVEQMYFQYGHWTE 834  
QY 619 CSBSCGGVQEAHVVSCLNKOTREPAENLCVTSRRPQLLKSCNLDPCPARWIGKWSPC 678  
Db 835 CSVTCTGTGIRROTAAHCI-KKGRGMVKATCFDPTOPNGRKQKCKEAKCPRWAGWEAC 893

```
QY 679 SUTCGVGLQTRVDFCSHLLSRENNETVILADELCQPKPSTVOACNRFNCPNPPAWYPAQWQ 738
Db 894 SATCG-----PH-----900
QY 739 PCSRTGGGVQKREVLCQKORMADGSLFELPETFCS-ASKPACQQAACKDD-CPSEWLLSD 796
Db 901 -----GEKKRTVICIQTWVSD-QUALPPDQCQLLKPXLLSCNRDILCPSDWTGVN 951
QY 797 WTECSTSCGEGQTRSAICRKLKTKGLSTVWVNSTLCPPLPFSSIRPCMILATCARPGRES 856
Db 952 WSECSVSCGGVIRSVCTAKNDE-----PCDVTREPNSREALG-----991
QY 857 TKHSPHIAARKVYIOTRRQRKLHFVVGGFAYLLPKTAVVILRCP-ARRVKKPLITWEKDG 915
Db 992 -----LQCFSSRRVLKP-----1004
QY 916 QHLISSTHVAPFGYKLIHRLKPSDAGVYTCAGPAREHFVILKLGGRKLVARPLSPR 975
Db 1005 -----1004
QY 976 SEEVLAGRKGPKKEALQTHKHQNGIFSGSKAEKRGLEAANPGSRYDDLVSRLLGQGW 1035
Db 1005 -----NKGITSGKNPPTLKPVPPTSR-----PRMLTTTGP 1037
QY 1036 GELLASWEAODSAERNWTSEE-DPGAEOVLLHLPFTWTEQRDLDDILGNLSQOPELRD 1094
Db 1038 ESMSTPTAISPTASPTASKEGLGKG-----WQDSSTQPE-----1074
QY 1095 LYSKHLVA--QLAEIFRSHLEHODVILKPSERRTSPVTLSPHKHVSFGFSSRTSTG 1152
Db 1075 LSSRYLISTGSTSPILTS-----QSLIQPSEENVSSDTGP-----TSEGILVATTISGS 1126
QY 1153 AGGSRPRPKP-----TILRKISAAQLSASEVWTHLGTVALASGTVLLHCEAIGH 1207
Db 1127 GLSSRNPIWVTPPTTLTK-----GPEMEHSG-----1157
QY 1208 PRPTISWARGNEVQSRDRILODDSLQILAPVEADVGYTCNATNALGYDSVIAVTL 1267
Db 1158 -----SGEERE-----QPEDK-----DESNPVITKIRVP-----1182
QY 1268 AGPLVKTSRMTINTEKPAVTVDIGSTIKTVQGVNVTINQVAGVPEAEVTFERNKSKL 1327
Db 1183 GNDAPVESTEMPL-----APLTPDPL-----SRESWPPPTST- 1214
QY 1328 GSPHLLHEGSLLLTVNSSDDQGLYSCRAANLHGLTESTQLLLDPPQVPTQLEDIRALL 1387
Db 1215 -----VMEGLL-----PSQRPTTSE-----1229
QY 1388 AATG-PNLPVSLTSPIGTQVLDPGNSALLGCPKIG--HPVP-----NITW 1430
Db 1230 --TGTPRVGNVTE-----KPANTLL--PLGSDHQPEPSGKTANRHLKLPNNNQ 1276
QY 1431 FHGQPTVATGLTHHILAAGQILQVANSLSGSGGEPSCLAQNEAGVLMOKASLVIQDYW 1490
Db 1277 TKSEPVLTEDAT-SLITTEGFLNLSNYKOLTHGSA-----H 1315
QY 1491 WSDVRLATCSAGNNGVQVQPRCLLNSTEVNPAHCAKGVKRPVAVQPI-ACNRRDCPSRW 1549
Db 1316 WIVGNWSECSTTCG-LGAYWKRVECT--TTQMD-SDCAAIQRC--DPAKRCHLRPC-AGW 1367
QY 1550 MVTWSACTSCGGGVOTRRVTC-----OKLK-----ASGISTPVSNDCQVAKRPV 1597
Db 1368 KVGWNSKSRNCSGGFKIREIQCVDSRDHRLNRFHFCQFLAGIPPLUS-----1415
QY 1598 DTQACNQLCVWEAFSSWGQNGPCIGPHLAVQHRQVFCQTRDGTITLFPSECCSALPRVPS 1657
Db 1416 --MSCNPEPCBAWQVWPSCSRSCGG--GVQERGVEF-----PGGLCDWTKRPTS 1462
QY 1658 TQNCWSERACSVHWRVSLWTLCATCGNYGFSRVECVHARTKAVPEH--LCSWGPRPA 1715
Db 1463 TMSCNELHC-CHWATGNWDLCTSCGG-GFQKRIVQCVPSSEGNKTEDQDQCLCHKPRPP 1520
QY 1716 NWQRCSNITPCE--NMBC---RDITRYCEKVKQLKLCQLSOFKSRCCGTC 1759
```

```
Db 1521 EFKKCQQACKKSADILLCTKDKLSASFQCLTKAMKKCSVPTVRAECFCSC 1570

RESULT 3
AT10_HUMAN STANDARD; PRT; 1077 AA.
ID AT10_HUMAN
AC Q9H324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type I repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -|- COPACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC similarity).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: AF163762; AAC35563.1;
DR InterPro; IPR002870; Rep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptidse.
DR Pfam; PF01562; Rep_M12B_propep; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF00090; tsp_1; 5.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; zinc; Glycoprotein; zymogen;
KW Repeat; Extracellular matrix.
FT CHAIN 1 207 BY SIMILARITY.
FT PROPEP <1 207 ADAMTS-10.
FT METAL 208 1077 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 366 366 ZINC (CATALYTIC).
FT METAL 367 367 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 SPACER.
FT DOMAIN 680 802 CYS-RICH.
FT DOMAIN 803 860 TSP TYPE-1 1.
FT DOMAIN 861 918 TSP TYPE-1 2.
FT DOMAIN 919 976 TSP TYPE-1 3.
FT DOMAIN 977 1031 TSP TYPE-1 4.
FT DOMAIN 1032 1031 TSP TYPE-1 5.
FT CARBOHYD 64 N-LINKED (GLCNAC. . .) (POTENTIAL).

```



FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	714	714	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	769	769	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	866	866	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1077	AA; 118072 MW; 3914DE18DCBFF587	CRC64;	

  

Query Match		8.4%;	Score 807;	DB 1;	Length 1077;									
Best Local Similarity		32.9%;	Pred. No. 4e-42;											
Matches	197;	Conservative	71;	Mismatches	224;	Indels	106;	Gaps						
QY	30	EDRGLDWP	36	SEKSRCTCGG	ASYSLLRC	-----	LSSKSC	EGNRIYRTCSNVDCPP	83					
Db	518	EGVDGAMP	217	TPWGDGSR	TCGGVSSSRHCD	SPRPTIGG	KYKCYCLGER	RHSRSCNTDCPP	577					
QY	84	EAGDFRAC	QCSA	HNVDVXH	GOFEWLP	VPVNDP	NPSCSLK	CAQAGTIV	VELAPKVLDTGR	143				
Db	578	GSQDFREV	QCSSEFDS	IFPRKFT	YKWKIYR	GGVYKAC	SLTSLA	EGNFYTER	AAAVVDGP	637				
QY	144	CYTESLDM	ICSLGCI	VGCDHQ	IGSTVK	KEDNC	GVNCGD	STCRLV	RGVYKSOLSATKSD	203				
Db	638	CRPTDVCV	SGECKH	GVGCDRL	VEDLREDK	RCVCGD	GSACETI	EGVF	SPASPGAGVE	696				
QY	204	TVVAIPG	SGSHI	-----	RLVLKGP	DHLYE	TKTLO	GTGEN	SLSTGTFLVDNS	253				
Db	697	DVVMIPK	GVSHIF	IQDNL	SLSHAL	KLG	-DQESL	LEGLPG	TQPQHLPLAGT	749				
QY	254	VDQKFPD	K-EILRM	AGPITAD	FIVKIR	NSGSAD	SVQVIF	YQPII	-----	HRWRETFD	306			
Db	750	FQLRQGD	QVQSLE	ALGPIN	ASLIV	WVL	-ARTEL	PALRYR	FNAPIARD	SLPPYSWHYAPW	808			
QY	307	FPGSATGG	QYQLTS	AECDYDL	RSNRV	VADQY	CHYYP	ENIKPK	PKLQE	-CNLDPC	PASDGY	365		
Db	809	TKCSAQ	CAGGSQ	VAECRN	QLDSS	AVAPHY	CSAHSK	---LPK	RQACNTE	FECP	-----	859		
QY	366	KQIMPV	DLVHPL	PRWATP	WTAC	SSCGG	IGIOS	AVSCVE	EDIQGH	VTISVEW	---KCM	421		
Db	860	-----	PDWV	GVNWSL	CRSCD	AGVRS	SVVC	-----	QRRV	SAAE	KALDD	902		
QY	422	YTPKMPI	AOPCN	IFDC	-PKWLA	QWSP	CTVTCG	QGLRV	WVLC	--IDIRG	-MHTGC	SPK	477	
Db	903	PQPRPV	LACHG	PTCP	PPWAL	MDSE	CPSCG	PLRHR	VJLCK	SADHR	ATLPPA	HCSPA	962	
QY	478	TKPHI	KECIV	TPCYK	PKLE	VPFA	KLPWF	FOAQ	ELEGA	AVSEPS	FIFPA	NSACTVT	537	
Db	963	AKPPAT	MRCNL	-----	RRCP	-----	-----	PAWV	AGE	CECSA	Q	992		
QY	538	CGVGTQ	VRVRC	QVLL	SFSQ	VADLP	IDBC	-BGP	KPAS	ORACY	AGPC	SEIPE	NPDE	594
Db	993	CGVGQ	RQSR	SVRC	-----	TSHTG	QA	SHECTE	ALRP	PTTQ	CEA	-KCDSP	TGDP	1042

  

RESULT	4
AT33	HUMAN
ID	AT33 HUMAN
AC	AT15072;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (fragment).
GN	ADAMTS3 OR KIAA0366.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid	9606;
RP	[1]
RN	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RX	MEDLINE=97349984; PubMed=9205841;
RA	Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

"Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";

RL DNA Res. 4:141-150(1997).

CC -I- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -I- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -I- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -I- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -I- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.

CC -I- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.

-----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[illegible]

Query Match	6.8%	Score 651;	DB 1;	Length 1201;
Best Local Similarity:	31.28			

QY 28 SEEDRDGLWDAGWPWSECSRTCGGGASYSLRRCLS-----SKSCEGRNTRYRTCSNVC 81



[illegible]

DR InterPro: IPR000884; TSP1.  
DR InterPro: IPR000130; Zn\_Mtpeptdse.  
DR Pfam: PF01562; Pep\_M12B\_propep; 1.  
DR Pfam: PF01421; Repolysin; 1.  
DR Pfam: PF00090; tsp\_1; 3.  
DR SMART: SM00209; TSP1; 3.  
DR PROSITE: PS0215; ADAM\_MEPRO; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
DR PROSITE: PS0092; TSP1; 2.  
DR PROSITE: PS00427; DISINTEGRIN\_L; FALSE NEG.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
Repeat: Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 54  
FT PROPEP 55 252  
FT CHAIN 253 967  
FT SITE 205 205  
FT METAL 401 401  
FT ACT\_SITE 402 402  
FT METAL 405 405  
FT METAL 411 411  
FT DOMAIN 476 518  
FT DOMAIN 559 615  
FT DOMAIN 616 724  
FT DOMAIN 725 837  
FT DOMAIN 858 907  
FT DOMAIN 908 967  
FT DOMAIN 194 198  
FT CARBOHYD 547 547  
FT CARBOHYD 720 720  
FT CARBOHYD 764 764  
FT CARBOHYD 782 782  
FT CARBOHYD 945 945  
FT CONFLICT 21 21  
FT CONFLICT 26 31  
FT CONFLICT 49 49  
FT CONFLICT 72 72  
FT CONFLICT 79 79  
FT CONFLICT 249 249  
FT CONFLICT 262 265  
FT CONFLICT 607 607  
FT CONFLICT 936 936  
FT CONFLICT 962 962  
SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;  
  
Query Match 6.4%; Score 613; DB 1; Length 967;  
Best Local Similarity 29.6%; Pred. No. 3.3e-30;  
Matches 145; Conservative 65; Mismatches 166; Indels 114; Gaps 19;  
  
QY 34 GLWDAMPWSECSRTCGGASYSLRRC-----LSSKSCGRNTRYRTCSNVDCPPAG- 86  
DB 560 GSWGPWGPWDCSNTCGGVQYTMRECDNPVKNGKGYCEGKRYRYSRNCIEDCPDNNGK 619  
QY 87 DFRQAQCSAHNDVKH-----HQGFVWLP--VSNPDNPCLKCAKGTLLVVELAPKVL 140  
DB 620 TFREOQCEAHNEFSKASFGNEPTWTPKYAGVPKORCKLTCEAKGIGYFFVLQPKVVD 679  
QY 141 GTRCYTESLDMCISGLCQIVGCDHQLGSTVKEDNCGVNGDSTCRILVRGQYKSQLSATK 200  
DB 680 GTPCSPDSTVCVQGVKACQDRIDSKKKFKDCGCVGGNGSTCKKISG---FVTSRTP 736  
QY 201 SDDTVAIPYGRHRLVLKGPDLHLYETKLTQTKGENSL-----SSTGTFLVDN----S 252  
DB 737 GYHDIVTIPAGATNIEVKHRNP-----RGRNRNGSFALRAADGTVLINGNFTLS 786  
QY 253 SVDFQKFDKELLKRMAGPLTADFTVKIRNSGADS--TVQFIYQPIIHRKRETDFFPCS 310  
DB 787 TLEQDLTKGIVLRVSGSSAA--LERTSFSPLEPTIQVLMV----- 828  
QY 311 ATCGGGYQLTSAECYDLRSNRWADQYCHYYPENIKPKLOENLDCPPASDGY----- 365  
DB 829 -----GHAL-----RPKIK-----YTFPMK 843

QY 366 KQIMPYDLYHPLRPWEATPWTACSSSGGSIOSRAVSCVEEDIQGHVTSVEENKCMVTPK 425  
DB 844 KKTPEPNAITPTSEWVIEWGECSKTCGGWQRRVVEEC--RDINGHPAS-----EKAKEVK 897  
QY 426 MPTAQPCNLFDCPKWLAQEWSPCTVTCGGRLRYRVVLICIDHRG--MHTGCSPTKTP-HI 482  
DB 898 PASTRCADLPFCPRWQVGDWSPCSKTCGKGYKRTKLCLSHDGVLSNESCPLKPKPHY 957  
QY 483 KEECIVPTPC 492  
DB 958 IDFCIL-TQC 966  
  
RESULT 7  
AT52\_HUMAN  
ID AT52\_HUMAN STANDARD; PRT; 1211 AA.  
AC O95450;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and  
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)  
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I  
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI)  
DE (Procollagen I/II amino-propeptide processing enzyme).  
GN ADAMTS2 OR PCNP1 OR PCPNI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI).  
RC TISSUE=Skin;  
RX MEDLINE=99347935; PubMed=10417273;  
RA Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,  
RA Wiertelicki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,  
RA Byers P.H., Lapiere C.M., Prockop D.J., Nusgens B.V.;  
RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis  
are caused by mutations in the procollagen I N-proteinase gene.";  
RL Am. J. Hum. Genet. 65:308-317(1999).  
CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR  
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO  
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN  
CC COLLAGEN BIOSYNTHESIS.  
CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
CC alpha-1(I) at Pro-I-Gln and of alpha-1(II) and alpha-2(I) chains  
CC at Ala-I-Gln.  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO  
CC COLLAGEN TYPE XIV (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;  
CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-  
CC PROCOLLAGEN PEPTIDASE ACTIVITY.  
CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON  
CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS  
CC SYNDROME TYPE VIIC (EDS-VIIC). A RECESSIVELY INHERITED DISORDER  
CC CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND  
CC BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY  
CC PROCESSED AT THE AMINO TERMINUS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.  
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: AJ003125; CRA05980.1; -  
 MEROPS; M12.301; -  
 MIM; 604539; -  
 MIM; 225410; -  
 InterPro; IPR001762; Disintegrin.  
 InterPro; IPR002870; Pep\_M12B\_propep.  
 InterPro; IPR001590; Reprolysin.  
 InterPro; IPR000884; TSPI.  
 InterPro; IPR000130; Zn\_MTpeptdse.  
 Pfam; PF01562; Pep\_M12B\_propep; 1.  
 Pfam; PF01421; Reprolysin; 1.  
 Pfam; PF00090; tsp\_1; 4.  
 SMART; SM00209; TSPI; 4.  
 PROSITE; PS0215; ADAM\_MEPRO; 1.  
 PROSITE; PS50092; TSPI; 1.  
 PROSITE; PS00142; ZINC\_PROTEASE; FALSE\_NEG.  
 PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;  
 Alternative splicing;  
 SIGNAL 1 29  
 PROPEP 30 253  
 CHAIN 254 1211  
 METAL 408 409  
 ACT\_SITE 409 409  
 METAL 412 412  
 METAL 418 418  
 DOMAIN 480 560  
 DOMAIN 561 617  
 DOMAIN 618 722  
 SITE 691 693  
 DOMAIN 723 851  
 DOMAIN 852 911  
 DOMAIN 912 974  
 DOMAIN 975 1030  
 DOMAIN 40 43  
 DOMAIN 185 188  
 CARBOHYD 112 112  
 CARBOHYD 251 251  
 CARBOHYD 949 949  
 CARBOHYD 993 993  
 CARBOHYD 1031 1031  
 CARBOHYD 1098 1098  
 CARBOHYD 1145 1145  
 CARBOHYD 1150 1150  
 VARSPLIC 544 566  
 VARSPLIC 567 1211  
 SEQUENCE 1211 AA; 134722 MW; BECEE25C23CAD2D CRC64;

Query Match 6.3%; Score 608.5; DB 1; Length 1211;  
 Best Local Similarity 22.9%; Pred. No. 8.6e-30;  
 Matches 183; Conservative 87; Mismatches 212; Indels 317; Gaps 29;

32 RDGLWDANGPWSECSRTCGGGASYSLRRC-----LSSKSCEGRNIYRTRCSNVDCPPEA 85  
 560 RDGSGWGAAPFGSCSRTCGTGKVFTRQCDNPHFANGRTCSGLAYDFQLCSRODCPDSL 619  
 86 GDFRAQOCS-----AINDVKHHGQFVEWLPVSN-DPDNPSCLKCAKGTTLVVELAPKV 138  
 620 ADFEEQCRQWDLYFEHGDQHH-----WLPHEHRAKERCHLYCESRETGETVVSMMKRV 674  
 139 LDGTRC-YTESLDMWCISGLCQIVGCDHDLGTSVKEQDNCVCGNDGSGTCLRLVROYKSQLS 197  
 675 HDGTRCSYKDAFSLCYRGDCRKVKGCDGVIGSSKQEDKCGVCGGNDHSHCKYKGTF--- 729

QY	198	ATKSD	-----TVVAIPYGSRRIRVLKGPDPHLYLETKLOGK-----GENSL-SSTGTF	247
Db	730	-TRSPK	HGYIKMFEIPAGARHLLIQEVDATSHLAVKNLETGTFILNEENDVDASKTF	788
QY	248	LVNDSSVDP	QKFPDKIELRMAGPLTADFIIVKRSNGSADSTVQFIYFOPIIHRWRETDF	307
Db	789	IANGWEY	RDDEGRETLOTGPLHGTTIVLIVFVG--DTRVS-LTYKYMIH-----838	
QY	308	PCSATCGGY	QLTSAECYDLSRNVVADYCHYYYPENIKPKKLOECNLDPCCASQYKQ	367
Db	839	-----	DSLNVDDNVLEDSVVY-----856	
QY	368	IMPYDLYHPL	PRWEAIPWTACSSCGGGISQRAVSCVEEDIQGHVTSVEWKCMYTPKM	427
Db	857	-----	EWALKKWSPCS KPGGSGQTKYC-----881	
QY	428	IAOPCNIFDC	PKWLAQWSPCTVTCGGGLRYRVVLCIDHRGMHTGCCPKTKPHKECI	487
Db	882	-----	RR-----LDHKMVRGFCALSKPKAIRAC	908
QY	488	VPTPCYKPEKLP	VEAKLPWFKQAELEGAASERPSFIPEAWSACTVTCG-VGTQVRI	546
Db	909	NPQEC	-----SQPVMVTGMEWPCSTCGRTGMQVRS	939
QY	547	VRCQVLLSFS	QSVADLPIDECGPKPQACYAGPCSGEIPFNDETGLGGLQDFD	606
Db	940	VRC	-----IQPLHD-----NTTSSVIAKHCNDARPE-----965	
QY	607	ELYDWEYEGFT	KCSGCGGVQVAVVSLNKKOTREPAENLVCVTSRRPPQLLKSCNLDP	666
Db	966	-----	RR-----ACSREL	975
QY	667	PARMEIGKWS	PCSLTCGVLQTRDVFCSHLLSREMNETVILADE---LCRPQKPSVQAC	723
Db	976	PGRWAGPWS	QCSVTCGNGTQRPVPCR-----TADDSFGICQEEPERTARTC	1023
QY	724	NRFNCP	-----PAWYPAQM-----OPCSRT-----CGGGVKREVLCKORMADGSF	764
Db	1024	RLGPCPRNI	SDPKSSVYQWLRSRPDPDSPIRKISSKGHCQG---DKSIFCR-----1072	
QY	765	LELPETPCSA	---SKPACQ	780
Db	1073	MEVLSRYCS	IPGYNKLSCK	1091
RESULT	8			
ID	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
AC	P97857	O54768;		
DT	30-MAY-2000	(Rel. 39, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	ADAMTS-1	precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).		
DE	GN			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata;			
OC	Mammalia; Euthera; Rodentia;			
OC	NCBI_TaxID=10090;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SVJ;			
RC	MEDLINE=98110583;			
RT	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RT	"The exon/intron organization and chromosomal mapping of the mouse			
RT	ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RL	Genomics 46:466-471(1997).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=97150761;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.;			
RA	Matsushima K.;			
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase			

disintegrin family protein with thrombospondin motifs as an inflammation associated gene.";  
J. Biol. Chem. 272:556-562(1997).  
[3]  
RX CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.  
MEDLINE-9303657; PubMed-10373500;  
Kuno K., Terashima Y., Matsushima K.;  
\*ADAMTS-1 is an active metalloproteinase associated with the extracellular matrix.";  
J. Biol. Chem. 274:18821-18826(1999).  
[4]  
RX FUNCTION.  
MEDLINE-20389568; PubMed-10930576;  
Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M., Ohno H., Matsushima K.;  
\*ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";  
FEBS Lett. 478:241-245(2000).  
[5]  
RX FUNCTION, AND INDUCTION.  
MEDLINE-20243757; PubMed-10781075;  
Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W., Richards J.S.;  
\*progesterone-regulated genes in the ovulation process: ADAMTS-1 and cathepsin L proteases.";  
Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).  
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE INHIBITOR ACTIVITY (BY INVOLVED IN ITS TURNOVER, HAS ANGIOGENIC INHIBITOR ACTIVITY (BY SIMILARITY). ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-|-LEU-1692 SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.  
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX.  
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA CELLS OF PREOVULATORY FOLLICLES.  
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 7.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AB001735; BAA24501.1; ALT\_INIT.  
DR EMBL; D67076; BRA11088.1; ALT\_FRAME.  
DR MEROPS; M12.222;  
DR MGD; MGI:109249; Adamts1.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR002870; Pep\_M12B-propep.  
DR InterPro; IPR001590; Reprolysin.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR000130; Zn\_Mtpeptidse.  
DR Pfam; PF01562; Pep\_M12B-propep; 1.  
DR Pfam; PF01421; Reprolysin; 1.  
DR Pfam; PF00090; tsp\_1; 3.  
DR SMART; SM00209; TSP1; 3.  
DR PROSITE; PS50215; ADAM\_MEPHO; 1.  
DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
DR PROSITE; PS50092; TSP1; 3.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 48  
FT PROPEP 49 253  
FT CHAIN 254 968  
FT SITE 206 206  
FT METAL 402 402  
FT ACT\_SITE 403 403  
FT METAL 406 406  
FT METAL 412 412  
FT DOMAIN 477 559  
FT DOMAIN 560 617  
FT DOMAIN 618 725  
FT DOMAIN 726 850  
FT DOMAIN 851 909  
FT DOMAIN 910 968  
FT DOMAIN 195 199  
FT CARBOHYD 548 548  
FT CARBOHYD 721 721  
FT CARBOHYD 765 765  
FT CARBOHYD 783 783  
FT CARBOHYD 946 946  
FT MUTAGEN 403 403  
FT CONFLICT 335 335  
FT CONFLICT 425 425  
FT SEQUENCE 968 AA; 105841 MW; 42E8DA55499FB6C1 CRC64;  
  
Query Match 6.2%; Score 597; DB 1; Length 968;  
Best Local Similarity 29.1%; Pred. No. 3.3e-29;  
Matches 143; Conservative 51; Mismatches 180; Indels 118; Gaps 16;  
  
QY 34 GLDWDGWPSCRTGGGASYSLRRC-----LSSKCEGRNRYRSCNVDCPEAG- 86  
DB 561 GSWGPWGPWGDSCRTGGGQVQYTRCDNPVPKNGKYGCEGRKYRSCNIEDCDPDNNGK 620  
QY 87 DFRQQCSAHNDVKH---HGQFYEWLP--VSNDPDNCSLKCAQGTLLVVELAPKVL 140  
DB 621 TFRFQCEAHNEFSKASFGNEPTVETPKYACVSPKDKCKLTCEAKGIGYFVLQPKVVD 680  
QY 141 GTRCVTESLDMCISGLCQIVGCDHQLGSTVKEDNGVCGNGDSTCLRVGRQYKQSLSATK 200  
DB 681 GTPCSPDSTSVQVQGVQYKAGCDRIIDSKKFKDKGCGVGGNGSTCKKMSG---IVTSTRP 737  
QY 201 SDDTVVAIPYGSRRHR-----LVLKGPDLHYLETKTLOGTKQENSLSSTG 245  
DB 738 GYHDIVTIPAGATNIEVKHNRQGRNRNGSFLAIRAADGTYLNGNFTLTLEQDLTYKG 797  
QY 246 TFLVDNSSVDQKFPDKELRMAGPLTADFTVTKIRNSGSA-DSTVQFIYQPIIHRWE 303  
DB 798 TVLRYSGSSAALER-----IRSFPLKEPLTIQVWGHALRPKIKFTYFM----- 843  
QY 304 TDFPPCSATCGGQYLTSAECYDLRSNRVADQYCHYYPENIKPKPKLQECNLDPCPSD 363  
DB 844 -----KKKTESFNAIP----- 854  
QY 364 GYKQIMPYDLYHPLRPWEATPTACSSCGGGLQSPAVSCVREDIOGHVTSVEENKCMYT 423  
DB 855 -----TFSEWVTEWGECSKTCGSGWRVRVQC--RDINGHPAS-----ECAKE 896  
QY 424 PKMPIAQCNIQFDCPKWLAQWSPCTVTGQGLRYRWVLCIDHRG--MHTGGCSPTKP- 480  
DB 897 VKPASTRPCADLPCHPWQVCDMSPCSKTCGKYKKTACVSHDGGVLSNESCDPLKKPK 956  
QY 481 HIKEECIVTPPC 492  
DB 957 HYIDFCTL-TQC 967  
  
RESULT 9  
ATSL\_HUMAN STANDARD; PRT; 967 AA.  
ID ATSL\_HUMAN

AC Q9UR18; Q9URP8; Q9URH3; Q9P2K0; Q9NSJ8;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).  
 GN ADAMTS1 OR METH1 OR KIAA1346.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;  
 RA "Cloning, characterization and mapping on human chromosome 21 of the  
 RA orthologue of murine Adamts-1.";  
 RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RL SEQUENCE FROM N.A., AND FUNCTION.  
 RN [3]  
 RN MEDLINE=99367466; PubMed=10438512;  
 RA Vazquez P., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,  
 RA Lombardo M., Iruela-Arispe M.L.;  
 RA "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new  
 RA family of proteins with angio-inhibitory activity.";  
 RL J. Biol. Chem. 274:23349-23357(1999).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
 RA Minoshima S., Shimizu N., Nordsiek G., Mitsuyama S., Antonarakis S.E.,  
 RA Schaefer M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
 RA Leirach H., Reinhardt R., Raspo M.-L.;  
 RA "The DNA sequence of human chromosome 21.";  
 RL Nature 405:311-319(2000).  
 RN [6]  
 RN SEQUENCE OF 418-967 FROM N.A.  
 RC TISSUE=Melanoma;  
 RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOLYCAN, AND MAY BE  
 CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR  
 CC ACTIVITY. ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH  
 CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER  
 CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.  
 CC -!- CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 1938-GLU-1-LEU-1939  
 CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.

CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF170084; AAF15317.1; -  
 DR EMBL; AF060152; AAD48080.1; ALT\_INIT.  
 DR EMBL; AF207664; AAF23772.1; -  
 DR EMBL; AB037767; BAA92584.1; ALT\_INIT.  
 DR EMBL; AP001697; BAA95502.1; -  
 DR EMBL; AL162080; CAB82413.1; -  
 DR MIM; 605174; -  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR002870; Pep\_M12B\_propep.  
 DR InterPro: IPR001590; Reprolysin.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR000130; Zn\_Mtpeptidse.  
 DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
 DR Pfam; PF01421; Reprolysin; 1.  
 DR Pfam; PF00090; tsp\_1; 3.  
 DR SMART; SM00209; TSP1; 3.  
 DR PROSITE; PS00215; ADAM\_MEPRO; 1.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE; PS50092; TSP1; 2.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Heparin-binding.  
 FT SIGNAL 1 49  
 FT PROPEP 50 252  
 FT CHAIN 253 967  
 FT SITE 198 198  
 FT METAL 401 401  
 FT ACT\_SITE 402 402  
 FT METAL 405 405  
 FT METAL 411 411  
 FT DOMAIN 476 559  
 FT DOMAIN 560 616  
 FT DOMAIN 617 724  
 FT DOMAIN 725 849  
 FT DOMAIN 850 908  
 FT DOMAIN 909 967  
 FT DOMAIN 843 846  
 FT CARBOHYD 547 547  
 FT CARBOHYD 720 720  
 FT CARBOHYD 764 764  
 FT CONFLICT 227 227  
 FT CONFLICT 468 468  
 FT CONFLICT 561 561  
 FT SEQUENCE 967 AA; 105383 MW; C189389324741ED1 CRC64;  
 SQ

Query Match 6.1%; Score 584.5; DB 1; Length 967;  
 Best Local Similarity 29.4%; Pred. No. 1.9e-28;  
 Matches 148; Conservative 59; Mismatches 168; Indels 129; Gaps 19;  
 QY 23 SRTARSEDR--DGLWDANGPWSCESTRCGGASYSLLRRC-----LSSKSGEGRNRYR 74  
 DB 547 NKIDRRKHEDTPFHGSGWGMGPDGDCSTCGGQVQYTMRECDNPVPKNGKCYCEGKRVYR 606

QY 75 TCSNVDCEPEAG--DRAQOCSAHNDVKH-----HGQFYEWLP--VSNDDPNPCSLKCOAQK 127  
Db 607 SCNLECPDNNKGTFRFEEQCEAHNEFSKASFGSPAVEMLPKYAGVSPKDRCKLIQCAQK 666  
QY 128 TTLVVELAPKVLIDGTRCYTESLDMCISGLCOIVGCDHQLGISTVKEDNGVCGNGDSTCRLL 187  
Db 667 IGYFFVLQPKVVDGTPCSPDSTSVCGQGVKACGDRIDSKKFKDKGVCVGGNGSTCKK 726  
QY 188 VGOYKQSLSAVKSDDTVAIPYKSGSHIRLVKGPDLHLYETKLTQGTGKENSLSL---SS 243  
Db 727 ISG---SVTSAPKPYHDIITIPGARNIE-----VKQRNQRSGNRNGSFLAIKAA 773  
QY 244 TGTFLVNDSSVDFQKPEKILRMAGPLATADFTVKIRNSGSDSTV-QTFYFQPIIHRWR 302  
Db 774 DGYI-----LNGDYIL-----STLEQDILMYKGVVLR-- 800  
QY 303 ETDFPPCSATCGGGYQLTSAECYDLNSNRVAD-----QYCHYYPENI 345  
Db 801 -----YSGSSAALERIRSFPLKEPLTIOVLTVGNALRPKIKYIFV--- 842  
QY 346 KPKPKLQECNLDPASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVAVSVE 405  
Db 843 --KKKESFNAIP-----TFSAWVIEWGECSKSCELGWORLVE-- 881  
QY 406 EDIQGHVTSVEEWKCMYTPKMPIAQCNIIDFCPKWLAQWSPCTVTTCGGLRVYRVLCID 465  
Db 882 RDINGQPAS---ECAKEVPASTRCPADHPCPQWOLGEWSSCKTCKGKYSKSLKCLS 937  
QY 466 HRG--MHTGCSPTKTP-HKKEC 486  
Db 938 HDGVLSHESCPLKPKHPIDFC 961

RESULT 10  
AT10\_MOUSE STANDARD; PRT; 450 AA.  
ID P58459;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-10 (EC 3.4.24.-) (A disintegrin and metalloproteinase with  
thrombospondin motifs 10) (ADAM-TS10) (Fragment).  
GN ADAMTS10.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Seldin M.F., Apte S.S.;  
RT "A novel gene of the ADAMTS family predicts ADAMTS-10, a  
metalloproteinase with unique structural features and expression  
pattern";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
MATRIX (BY SIMILARITY).  
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY  
similarity).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -1- SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE-1 DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF302012; AAK97226.1; -  
CC PROSITE; P550092; TSP1; 1.  
CC

KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Repeat;  
KW Extracellular matrix.  
FT NON\_TER 1  
FT DOMAIN <1 51 CYS-RICH.  
FT DOMAIN 52 174 SPACER.  
FT DOMAIN 171 232 TSP TYPE-1 1.  
FT DOMAIN 234 290 TSP TYPE-1 2.  
FT DOMAIN 294 348 TSP TYPE-1 3.  
FT DOMAIN 353 399 TSP TYPE-1 4.  
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 450 AA; 48861 MW; 2377D0E2CFBFECA CRC64;  
Query Match 5.9%; Score 564.5; DB 1; Length 450;  
Best Local Similarity 30.9%; Pred. No. 1.2e-27;  
Matches 151; Conservative 60; Mismatches 176; Indels 101; Gaps 21;  
QY 135 APKVLDTGRCTESLDMCISGLCOIVGCDHQLGISTVKEDNGVCGNGDSTCRLLYRGQYKS 194  
Db 1 AAAYVDTPCRPDIVDICVSGECKHVGCDRLKRDCKRCVCGGDSACETIEGVSP 60  
QY 195 QLSATKSDDTVAIPYGSRIH-----RLVLKGPDLHLYETKLTQGTGKENSLSST 244  
Db 61 ALPGTGIED-VVWIPKGSVHFIQDLNLSLHSLKLG-DOESLLEGLPGTPQPHRPLA 118  
QY 245 GTFLVNDSSVDFQKPEDK-EILRMAGPLATADFTVKIRNSGSDSTVQFIYFOPII----- 298  
Db 119 GT-----TFHLRQGPDAQSLEALGPINASLIIMVLAQAELPA-LHYRFNAPIARDALP 171  
QY 299 -HRWRETDFFPCSATCGGGYQLTSAECYDLNSNRVADQYCHYYPENIKPKPKLQE-CNL 356  
Db 172 PYSWHYAPWTKCSAQAGGQVQVVECRNQLDSSAVAPHYCSGHSK-----LPKQRACNT 227  
QY 357 DPCPASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVAVSVEEDIQGHVTSVE 416  
Db 228 EPCP-----PDWVVGNSCRSCDAGVRSRVVC-----QRRVSAE 265  
QY 417 EW---KCMYTPKMPIAQCNIIDFC-PKWLAQWSPCTVTTCGGLRVYRVLC--IDHRG- 468  
Db 266 EKALDDSSACQPRPVPVLEACQGPCPEWATLWDSECTPCGPGLRHVRVLCSDAQDST 325  
QY 469 MHTGCSPTKTPHIKEECIVPTPCYKPKELPVEAKLPWFKQAQLEEGAAVSEEPSFIP 528  
Db 326 LPPGCHLPAAPKPPSTMRCNL-----RRCP-----PARWVT 355  
QY 529 EAWSACTVTGCGTQVTRVRCQVLLSFQSVADLPIDEC-EGKPKPASQACVAGPCSGEI 587  
Db 356 SEWGECSQCGLGQQQRTVRC-----TSHTGQPSRECTEALRPSMQQCEA-KCDSVV 407  
QY 588 PEFN-PDE 594  
Db 408 PPGDGPEE 415  
RESULT 11  
AT56\_HUMAN STANDARD; PRT; 860 AA.  
ID AT56\_HUMAN  
AC Q9UKP5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-6 precursor (hc 3.4.24.-) (A disintegrin and metalloproteinase  
with thrombospondin motifs 6) (ADAM-TS 6) (ADAM-TS6).  
GN ADAMTS6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE-99395124; PubMed-10464288;  
RX



RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;  
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of  
 zinc metalloproteases.";  
 RL J. Biol. Chem. 274:25555-25563(1999).  
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY  
 CC DETECTABLE IN A NUMBER OF OTHER TISSUES.  
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF140674; AAD56357.1; -  
 DR HSSP; P34179; 1IAG.  
 DR MIM; 605008;  
 DR InterPro; IPR001762; Disintegrin.  
 DR InterPro; IPR002870; Pep\_M12B\_propep.  
 DR InterPro; IPR001590; Reprolysin.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR000130; Zn\_M12Bpropep.  
 DR Pfam; PF01562; Pep\_M12B-propep; 1.  
 DR Pfam; PF01421; Reprolysin; 1.  
 DR Pfam; PF00900; tsp\_1; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS0215; ADAM\_MEPRO; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR Repeat; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Hydrolyase; Extracellular matrix.  
 FT SIGNAL 1 21  
 FT PROPEP 22 244  
 FT CHAIN 245 860  
 FT METAL 403 404  
 FT ACT\_SITE 404 404  
 FT METAL 407 407  
 FT METAL 413 413  
 FT DOMAIN 453 509  
 FT DOMAIN 510 566  
 FT DOMAIN 567 668  
 FT DOMAIN 669 795  
 FT DOMAIN 796 852  
 FT DOMAIN 68 71  
 FT DOMAIN 662 665  
 FT CARBOHYD 99 99  
 FT CARBOHYD 172 172  
 FT CARBOHYD 222 222  
 FT CARBOHYD 234 234  
 FT CARBOHYD 676 676  
 FT CARBOHYD 843 843  
 FT CARBOHYD 97098 MW; E57213015DECB2C5 CRC64;  
 SQ SEQUENCE 860 AA; 5.3%; Score 506; DB 1; Length 860;  
 Query Match  
 Best Local Similarity 36.1%; Pred. No. 1.2e-23;  
 Matches 117; Conservative 40; Mismatches 105; Indels 62; Gaps 9;

QY 33 DGLWDAMPWSESRVTCGGGASVLRCLSL-----SKSCGGRNRYRTGCSNVDPEAG 86  
 DB 510 DGGWGPWSLWGECSRTCTGGGSSLRHCDSPAPSGGSKYCLGERKRYRSCTDPCPLGSR 569

QY 87 DFRAGQCSAHNDYKHGQFYEMLPVSNPDNPCLKQAKGTLVLVLAELKVLDTGRCYT 146  
 DB 570 DFEKQCADNDNPFRRGKYNWKPITGGGVKPCALNCLAEYNYFYTERAPAVIDGTQCN 629  
 QY 147 ESDMCISGLCQIVGDDHQLGTSVKEDNCVCNGDSTCLVRGQYKSQLSATKSDTIV 206  
 DB 630 DSLDICEGCKHVGCDNLGSDAREDCRCVCGGSGTCDIEGFFNDSLPRGYME-VV 688  
 QY 207 AIPGSRHRLVLKGDHLYLETKTQGTGKENSLSGTG-----TFLVDN 251  
 DB 689 QIPRGSVHI-----EVREAVMSKNYIALKSEGDYVINGAWTIDWPKFDVAG 736  
 QY 252 SSVDFOKEFPDK-ELLRMAGPLTADFI-----VKRNSGSASDTVOQIF 293  
 DB 737 TAFHYKRPDEPSLALGPTSENILVMVLQNLGIRYKFNVPITRTGSGDNEVGFTW 796  
 QY 294 -YOPIIHRWRETDFFPCSATCGG 316  
 DB 797 NHQP-----WSE-----CSATCAGG 811  
 RESULT 12  
 ID AT58\_HUMAN STANDARD; PRT; 890 AA.  
 AC Q9UPF79; Q9NZS0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-8 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS) (METH-2)  
 DE (METH-8).  
 GN ADAMTS8 OR METH2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=99367466; PubMed=10438512;  
 RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,  
 RA Lombardo M., Iruela-Arispe M.L.;  
 RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new  
 RT family of proteins with angio-inhibitory activity.";  
 RT J. Biol. Chem. 274:23349-23357(1999).  
 RN [2]  
 SEQUENCE OF 195-440 FROM N.A.  
 RX MEDLINE=20079168; PubMed=10610729;  
 RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;  
 RA "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on  
 RT mouse chromosome 9 and human chromosome 11.";  
 RL Genomics 62:312-315(1999).  
 CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.  
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER  
 CC EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND  
 CC KIDNEY.  
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC EMBL; AF060153; RAD48081.1; -
CC EMBL; AF175283; AAF25806.1; -
CC HSSP; P34179; ILAG.
DR MIM; 605175; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00209; TSPl; 2.
DR PROSITE; PS00215; ADAM_MPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 214 BY SIMILARITY.
FT CHAIN 215 890 ADAMTS-8.
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 365 365 BY SIMILARITY.
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 439 526 DISINTEGRIN-LIKE.
FT DOMAIN 527 583 TSP TYPE-1 1.
FT DOMAIN 584 690 CYR-RICH.
FT DOMAIN 691 832 SPACER.
FT DOMAIN 833 890 TSP TYPE-1 2.
FT DOMAIN 202 205 POLY-PRO.
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 195 195 E -> R (IN REF. 2).
FT CONFLICT 413 440 YLTLLGGHGDGCLLDAPGAALPPTGL -> FSGCHLQGW
FT HFKYLCVKVSELKCDLMP (IN REF. 2).
SQ SEQUENCE 890 AA; 96671 MW; 57D70E03D5739D3 CRC64;

Query Match 5.2%; Score 495.5; DB 1; Length 890;
Best Local Similarity 32.4%; Pred. No. 5.4e-23;
Matches 124; Conservative 49; Mismatches 135; Indels 75; Gaps 15;

QY 33 DGLWDAGPWSRCSRTCGGAGSYSLRCL-----SSKSGRNIYRFTCSNVDCPPEAG 86
DB 527 DSGWAPGPGWCSRTCGGAGVQFSHRECKDPEPQNGRYCLGRRAKYQSCHTECEPPDGK 586
QY 87 DPAQOC-----SAHNDVKKHGYEWLP--VSDNDPNPCSLKCAQKTTLLVELAPKVLVG 141
DB 587 SFEQOCCKYNAINTYDMDGNLLQWPKYAGVSPDRCKLFCRARGSEFKVFEEAKVIDG 646
QY 142 TRCYTESLDMCISGLCOIQVCDHGLSTGVKEDNCVGCNGDSTGLRVLRGYKSQLSATKS 201
DB 647 TLLGPTTALICVRGQCKAGCDHVDVSPRKLDRKCGVCGGKSGKRSKVSQ-----SLTPT 700
QY 202 D---DPAWAIYSGSRHI-----RLVLKGPDLHLYETKTLQGTGKGENSLSS 243
DB 701 NYGINDIVIPAGATNIDVQKQSHPGVQNDGNYLALKTADGGYLLNGNLAIISAIEQDILV 760
QY 244 TGTFLVNDSSV-----DFQKPPDK---EILRMAG-----PLIADFIVKIR 280
DB 761 KGTILKYSGIATLERLQSFRLPELTVQLLTVPGEVFPFKVKYTFVFPNDVDFSMQ-- 818
QY 281 NGSASDSTVQFIYQPLIH--RWRETDFFCSATFCGGYGLTSAECYDLRNRVVADQYCH 339
DB 819 -SSKERATTNII--QPLLHAQWVLGWSGCSSTCGAGWQRVTECRDPSCQ---ASATCN 872
```

```
QY 340 --YYPENIKPKPKLOECNLDPCP 360
DB 873 KALPEDAKP-----CESQLCP 889

RESULT 13
ATS8_MOUSE
ID ATS8_MOUSE STANDARD; PRT; 905 AA.
AC P57110;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).
GN ADAMTS8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RT mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART
CC AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

FT	DOMAIN	599	705	CYS-RICH.
FT	DOMAIN	706	847	SPACER.
FT	DOMAIN	848	905	TSP TYPE-1.2.
FT	CARBOHYD	415	415	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	480	480	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	506	506	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	905 AA;	98879 MW;	124D4132B33A0CAE CRC64;

  

Query Match		4.9%	Score 474.5;	DB 1;	Length 905;
Best Local Similarity		30.9%	Pred. No. 1.1e-21;		
Matches 117;		Conservative	53;	Mismatches 142;	Indels 67;
Qy	33 DGLWDAMPSECSTCGGASYSLRRC-----LSSKSCEGNRIRYRTCSNVDCSP	86	601		
Db	542 DGDGWPMPWCQCSRTCGGGIQFSNRECDNPMQPONGRFLGERVXYQSCNTECPNGK	601	601		
Qy	87 DFRACQCSAHNDVKH---HCQFWEHLPVN-DFDNPCLSKQAKGTTLVVELAPKVL	141	141		
Db	602 SFREQCEKYNAYNHTDLGNFLQWPKYSGVSPDRCKLFCRARGSEKFKVEAKVID	661	661		
Qy	142 TRCYTSLDMCISGLCOIVCGHOLGTSVKEDGCGVNCGDGSTRCLVRGQYKSQLSATKS	201	201		
Db	662 TLCGPDLTSLICVRQCQKACGDHVNVPKPKLDKCGVGKGTACRISGSF-TPF	720	720		
Qy	202 DDTVAIPYSGRHRLVKG-----PDHLYLETKTLQG--TKGENSLSS-----TGT	246	246		
Db	721 D-IVTIPAGATNIDVKORSHPGVRNDGVSALAKTANGQYLLNGNLAI	778	778		
Qy	247 FLVNDSSV-----DFQKFPDK---EILRMAG-----EILRMAG-----PLTADFIVKIRNSG	283	283		
Db	779 ILKYSGSMTALERLQSFQALPELTVLQITVSGEVPFPKRVYTFVFPNDMDFVQNSKER	838	838		
Qy	284 SADSTQVFIQYLIHWRTEDFPCSACTGGGYQLTSAECYDLNSRNVADQYCH--YY	341	341		
Db	839 ATTNIIQSL--PSA-EWVLGDWSECPSTCRGSMQWRTVECDPSGQ---ASDTCD	891	891		
Qy	342 PENIKPKPKLQECNLDPCP	360	360		
Db	892 PEDAKP-----CGSQPCP	904	904		

  

RESULT 14	
ATSS_MOUSE	STANDARD; PRT; 930 AA.
ID	ATSS_MOUSE
AC	Q9R001;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2) (ADMP-2) (Implantin).
GN	ADAMTS5.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RP	[1]
RN	SEQUENCE FROM N.A.
RR	MEDLINE=93395124; PubMed=10464288;
RX	Herskain T.L., Hirohata S., Seidin M.F., Apté S.S.;
RT	"ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases.";
RL	J. Biol. Chem. 274:25555-25563(1999).
CC	!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGECAN IN ARTHRIC DISEASES. MAY PLAY A ROLE IN PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC	!- CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 392-GLU- ALA-393 SITE.
CC	!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC	!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR

INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE  
DESTRUCTION OF AGGREGAN IN ARTHRIC DISEASES. MAY PLAY A ROLE IN  
PROEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.  
-!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 392-GIU-|-ALA-393  
SITE.  
-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
-!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
MATRIX (BY SIMILARITY).  
-!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA  
BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO  
CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,  
CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN  
ARTHRITIC PATIENT.  
-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
-!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY  
SIMILARITY).  
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
-!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
-!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
EMBL; AF142099; AADA9577.1; -;  
EMBL; AP001698; BAA95504.1; -;  
EMBL; AP001697; BAA95503.1; -;  
EMBL; AF141293; AAF02493.1; -;  
HSSP; Q9PW35; 1BUD.  
MIM; 605007; -;  
InterPro; IPR001762; Disintegrin.  
InterPro; IPR001590; Reprolysin.  
InterPro; IPR000884; TSP1.  
InterPro; IPR000130; Zn\_M1peptdse.  
Pfam; PF01421; Reprolysin; 1.  
Pfam; PF00090; tsp1; 2.  
SMART; SM00209; TSP1; 2.  
PROSITE; PS50215; ADAM\_MEPRO; 1.  
PROSITE; PS00142; ZINC\_PROTEASE; 1.  
PROSITE; PS50092; TSP1; 1.  
PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix.  
FT SIGNAL 1 16 POTENTIAL.  
FT PROPEP 17 261 POTENTIAL.  
FT CHAIN 262 930 ADAMTS-5.  
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).  
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 411 411 BY SIMILARITY.  
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DOMAIN 485 566 DISINTEGRIN-LIKE.  
FT DOMAIN 567 623 TSP TYPE-1 1.  
FT DOMAIN 624 731 CYS-RICH.  
FT DOMAIN 732 874 SPACER.  
FT DOMAIN 875 930 TSP TYPE-1 2.  
FT DOMAIN 37 41 POLY-ALA.  
FT DOMAIN 257 261 POLY-ARG.  
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 136 136 A -> G (IN REF. 2).  
FT CONFLICT 614 614 R -> H (IN REF. 3).  
FT CONFLICT 692 692 P -> L (IN REF. 2).  
FT SEQUENCE 930 AA; 101715 MW; B64281502F28193B CRC64;

```
Query Match      4.8%; Score 462.5; DB 1; Length 930;
Best Local Similarity 29.6%; Pred. No. 6.3e-21;
Matches 118; Conservative 50; Mismatches 139; Indels 91; Gaps 15;

QY 34 GLWDAMPWSECSRTCCGGASYSLRRC-----LSSKSCSGRNIRYRTCSNVDCPPPEAGD 87
Db 568 GNMGSWGSQCSRSQGGVQFAYRHCNNPAPRNNGRYCTGKRAIYRSCSLMPCPPNGKS 627
QY 88 ERAOCCSAHN-----DYKHGQFYEWLP--VSNDDPNPCSLKCOAKGTTLLVVELAPKVLGD 141
Db 628 FRHEQCEAKNGYQSDAKGVKTFVEWVPKYAGVLPADVCKLTCAKCTGYVVFSPKVTDG 687
QY 142 TRCYTESLDMCISGLCQIVGCDHQLGSTVKEDNCGVNGDSTCRLVRGQYKSQLSATKS 201
Db 688 TECRPYSNSVCYRGKCVRTGCDGLIGSKLQYDKCGVCGGDNSSCTKIVGTENKK---SKG 744
QY 202 DTVVAIPYGSRIHL-VLKGPDH-----LYLETKITLQG---TKGENSLSTGTFLVDNSS 253
Db 745 YTDVVRPEGATHIKVRQFKAKDQTRFTAYLALKKNGEYLINGKYMISETIIDIINGT 804
QY 254 VDFQKFPDKELLRMAGPLTADIVKIRNSGSADSTVQFTFYOPITHRWRETDFPCCSATC 313
Db 805 V-----MNYSG-----WSHRDDF-----LH 819
QY 314 GGGYQLTSAECYDLRSNRVAD-----OYCHYYPENIKPKKLOECNLDPCPASDGY 365
Db 820 GMGYSATK-----EILIVQILATDPTKPLDVRYSEFFVPK--KSTPKVNSV-----TSHGS 867
QY 366 KOIMPYDLYHPLRWEATPWTACSSCGGGIOSRAVSC 403
Db 868 NKV---GSHTSQPQWVTGPWLACSRCTCDTGWHTRTVQC 902
```

Search completed: July 24, 2002, 04:28:19  
Job time: 352 sec



Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb.pat:AX224809	+	3956.50	2540.98	2.0e-133	5076
gb.pat:AX235356	+	3953.50	2452.98	3.6e-133	7260
gb.pat:AX224823	+	3762.00	2421.36	1.7e-126	4854
gb.pat:AX224825	+	3673.00	2359.34	4.7e-123	8578
gb.pr:AF176313	+	2910.00	1981.11	2.0e-96	1803
gb.pat:AX224819	+	2809.50	1913.81	1.1e-92	2538
gb.pr:AF237652	+	2625.50	1695.57	1.4e-86	2523
gb.pat:AX224821	+	2615.50	1689.82	9.2e-86	2316
gb.pat:AX235354	+	2597.50	1670.38	1.1e-84	6294
gb.pat:AX235410	+	2597.50	1670.37	1.1e-84	6303
gb.pr:AF251058	+	2558.00	1655.68	7.3e-84	1616
gb.pat:AX224815	+	2425.50	1568.17	5.5e-79	2175
gb.pat:AX149471	+	2346.50	1415.09	5.0e-76	2912
gb.pat:AX224817	+	2231.50	1444.30	4.4e-72	1953
gb.pr:AF251423	+	2041.00	1295.71	3.2e-63	54193
gb.pr:AC011701	+	2041.00	1285.23	3.2e-63	205510
gb.pat:AX235377	+	1629.00	1049.42	4.3e-50	5107
gb.pat:AX235379	+	1629.00	1049.42	4.3e-50	5107
gb.pat:AX235411	+	1629.00	1049.42	4.3e-50	5107
gb.pr:AX235359	+	1629.00	1049.42	4.3e-50	5107
gb.pat:AX235413	+	1625.00	1046.85	6.0e-50	5107
gb.pat:AX224811	+	1535.00	993.07	5.9e-47	1341
gb.pat:AX224813	+	1321.00	869.78	4.4e-40	1119
gb.in:AF061825	+	1280.00	827.80	9.5e-38	3604
gb.in:AF314171	+	1052.50	680.22	1.6e-29	4271
gb.pat:AX319857	+	1004.00	646.62	1.2e-27	5808
gb.pat:AX319858	+	999.00	645.38	1.1e-27	5808
gb.pr:AB037733	+	975.00	628.94	1.1e-26	4518
gb.in:AF205357	+	944.00	606.47	2.0e-25	7100
gb.pr:HS250725	+	921.50	595.39	8.4e-25	4614
gb.pat:ES5265	+	887.00	576.90	9.0e-24	5115
gb.pat:AX128433	+	886.00	571.49	1.6e-23	5061
gb.in:AE003765	+	885.00	540.89	1.8e-23	5353
gb.htg:AC014412	+	884.00	548.25	3.6e-22	86446
gb.in:AC009213	+	883.50	542.05	7.9e-22	182525
gb.in:AF043121	+	877.50	566.17	3.6e-23	5199
gb.in:AF078161	+	863.00	550.95	2.5e-22	11008
gb.pr:AF140401	+	859.50	555.45	1.4e-22	4550
gb.pr:AF069698	+	859.50	554.98	1.5e-22	4950

540 ACTGGATGGAACCTCGTTGCCAACACGCAGTCCTTGGACAATGTGTATCAGTG 589

155 lYLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrVallys 171  
: : : : :  
: : : : :  
590 GCATCTGTGAGCAGTGGGCTGGATCGCAACTGGGAAGCAATGCCAAG 639

172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188  
: : : : :  
: : : : :  
640 GAGCACAACTGTGAGATCTGTGGCGCGATGGCTCCACCCTGCAGCGTGT 689

188 lArqGlyGlnTyrlYsserGlnLeuSerAlaThrLysSerAspThrV 205  
: : : : :  
: : : : :  
690 ACGGGGAAATCAAAGTCACACAGTTTCTCTGAAAAAAGAGAGAATAAG 739

205 alValAlaIleProTyrgLyserArgHisIleArgLeuValLeuLysGly 221  
: : : : :  
: : : : :  
740 TAATTGCTGTCTCTTTGGGAAGTCGAAGTGTGAGAAATTACAGTGAAGA 789

222 ProAspHisLeuTyrlLeuGluThrLysThrLeuGlnGlyThrLysGly 238  
||| ||| : : : : :  
||| ||| : : : : :  
790 CTTGCCACCTCTTTATTAAGTCAAAAACACTTCAAGGAAGCANAGAGA 839

238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerVal 255  
: : : : :  
: : : : :  
840 ACACAGCTTTAAACAGCCCCGGCTTCTGTCGTAGAAAAACACACAGTGG 889

255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaglyProLeu 271  
: : : : :  
: : : : :  
890 AATTTCAGAGGGGCTCCGAGAGCAAACTTTAAGATTCCAGGACCTCTG 939

272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288  
: : : : :  
: : : : :  
940 ATGGCTGATTTCACTTCAAGACCAGGTACACTGCGCCCAAGACAGCGT 989

288 rValGlnPheIlePheTyrgLnProIleIleHisArgTrpArgGluThra 305  
: : : : :  
: : : : :  
990 GOTTCAAGTCTCTTTTACCAGCCCATTCACTCATCTAGTGGAGACAAC 1039

305 spPhePheProCysSerAlaThrCysGlyGlyGlyTyrgLnLeuThrSer 321  
: : : : :  
: : : : :  
1040 ACTCTTTTCCCTGCATCTGCAGCTGGAGAGGTTATCAGTCAATTCT 1089

322 AlaGluCystyrAspLeuArgSerAsnArgValAlaAspGlnTyrcy 338  
: : : : :  
: : : : :  
1090 GCTGAATGTGTGGATATCCCGCTTGAAGAGGGTAGTTCCTGACCATTAT 1139

338 shIfstyrrProGluAsnIleLysProLysProLysLeuGlnGluCysA 355  
: : : : :  
: : : : :  
1140 TCACITACTACCTCTGAAATGTAAACCAAAACCAAACTCGAAGAATGCA 1189

355 snLeuAspProCysProAlaSerAspGlyTyrlYsGlnIleMetProTy 371  
: : : : :  
: : : : :  
1190 GCATGGATCCCTGCCCATCAAGTCATGGATTAAGAGAGATAATGCCTAT 1239

372 AspLeuTyrlHisProLeuProArgTrpGluAlaThrProIlePhrAlaCy 388  
: : : : :  
: : : : :  
1240 GACCACTTCCAACCTCTCCFCGTGGGAACAATACTCTTGGACTGCATG 1289

388 sSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerCysValG 405  
: : : : :  
: : : : :  
1290 TTCCGTGTCTCTGGAGAGGGATTCAAGACGGAGCTTTTGTGTGTAG 1339

405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421  
: : : : :  
: : : : :  
1340 AGGAATCCATGCATGGAGAGATATTGACAGTGGGAAGTAATGGAAGTGCATG 1389

422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438  
: : : : :  
: : : : :  
1390 TAGGACCCCAAMCCCAAGGTTATGCAAACTGTAAATCTGTTGATGCC 1439

438 olYstrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455  
: : : : :  
: : : : :  
1440 CAAGTGGATTGCCATGGAGTGGTGTCTCACTGCACAGTCACTTGTGCCGAG 1489

455 lyleuArgTyrArgValValLeuCysIleAspHisArgGlyMethIsthr 471  
1490 GGTACGGTACCGGGTGTTCGTGTATTAACCAACCGGGAGACGATCAT  
472 GlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIleVa 488  
1540 GGGGGTGCATTCACCAACCACTCAAGTTCACATCAAGAAGAATGTGTCA 1589  
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505  
1590 TCCCATCCCGTGTATTAACCAAAAGAAAGTCCAGTGGAAAGCAAAAT 1639  
505 euProTrpPheLysGlnAlaGlnLeuGluGlyAlaAlaValSer 521  
1640 TGCCTTGGCTGAACAACAGCAACAAGACTAGAAAGACAGCAATATCAACA 1689  
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538  
1690 GAAGAACCACAGTTCATTCAGAACCCCTGGTCAGCTCCAGTACCACGG 1739  
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555  
1740 TGGGCCAGGTGTGCAGTCCGCGAGGTCAAGTGCCTGTGTCTCCATCAT 1789  
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluClyProLys 571  
1790 TCACGACAGCTAGACTGAGCTGCCGAGGAGAGTGTGAAGGCCCAAG 1839  
572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProCy 583  
1840 CTGCCCAACGAACGGCCCTGCCTCTGGAAGCATGTGATGAGACCGCGGC 1889  
583 sSerGlyGluLeuProGluPheAsnProAspGluThrAspGlyLeuPheG 600  
1890 CTCCGAGAGCTAGACATCCCTCTCCCT..... 1917  
600 lyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPhe 616  
1918 .....GAGCACAGTGAGACGACTTACGACTGGAGTACGCTGGGTTC 1959  
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValse 633  
1960 ACCCTTGCACAGCAACATGCTGGGAGGCCATCAAGAAGCCATGACAGT 2009  
633 rCysLeuAsnLysGlnThrArgGluProAlaGluAsnLeuCysValT 650  
2010 GTGCTTACATATCAGACCCAGCACAGTCAATGACACTGTGTGTGATA 2059  
650 hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys 666  
2060 TGGTCCACCGTCTCCAGCCATGAGCCAGGCGCTGTAAACAGAGCCCTGT 2109  
667 ProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysGl 683  
2110 CCCCCAGGTGCATGTGGGTCTTGGGGCCCTGCTCAGTACCTGTGTGG 2159  
683 yValcylLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArg 700  
2160 AGTTGGAAATCAGACCCGAGATGTACTGCTCCCTGCAC..... 2196  
700 luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716  
2197 ..CCAGGGGAGACCCCTGCCCTCTCTGTAGAG..TGGCGAGATGAAAG 2241  
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733  
2242 CCCCATGCTTTACAAAGCATGCATGTGACTGCCCTCCCTGCGTGGCA 2291  
733 rProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyGlyValGlnL 750  
2292 CATTTGAAGATTCGACAGCTGTTCAGGACTGTGGCGGGGAATCAGA 2341









800 ysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArg 816  
2627 GTTCTGTGAGTTGTGTGGAAATCCAGAGAAGAGGAGGTGTCTCAA 2676  
817 LysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysPr 833  
2677 AGCTGGCAGCCAAAGTGGCGCATCCCTCAGTGAGATGATGTGCAG 2726  
833 oProLeuProPheSerSerSerIleArgProCysMetLeuAlaThrCysA 850  
2727 GGATCTACCAAGGTTCCCTTGTGAAGATCTGCCAGATGCCCTGAGTGA 2776  
850 laArg...ProGlyArgProSerThrLys.....HisSerProHis 862  
2777 GTPAAATCAATCAGACATCAACACAAAACCTTGTCGACGAGGTCCGAC 2826  
863 IleAlaAlaAlaArgLysValTyrIleGlnThrArgArgGlnArgLysLe 879  
2827 ATCTCAGTGTCCAGAGAGTCTATCATTCAGACAAGGGAAGAGAGCGTAT 2876  
879 uHisPheValValGlyGlyPheAlaTyrLeuLeuProLysThrAlaVal 896  
2877 TAACCTGACCATTTGGTAGCAGCGCTATTGTGCCCAACACATCCGTGA 2926  
896 allLeuArgCysProAlaAArgValArgLysProLeuIleThrTrpGlu 912  
2927 TTATTAGTCCCGTCGCAGGATTCAGAAATCTCTGATCCAGTGGGAG 2976  
913 LysAspGlyGlnHisLeuIleSerSerThrHisValThrValAlaLaph 929  
2977 AAGATGGCGTTGCCCTGCAGAACTCCAAACGGCTGGCATCACCAAGTC 3026  
929 eGlyTyrLeuLysIleHisArgLeuLysProSerAspAlaGlyValTyr 946  
3027 AGGCTCAATAAATCCACGCTCTTGTGCCCCCGACATCGCGTGTACC 3076  
946 hrCysSerAlaGlyProAlaAArgGluHisPheValIleLysLeuIleGly 962  
3077 GGTCAATTGAGGCTCTGCACAGGAAACAGTTGTGCTCAAGCTCAITGGT 3126  
963 GlyAsnArgLysLeuValAlaAArgProLeuSerProArgSerGluGlu 979  
3127 ACTGACACCGGCTCATCGACGCCCA..... 3153  
979 uValIleAlaGlyArgLysGlyProLysGluAlaLeuGlnThrHisL 996  
3154 .....GCCCTCAGGAGCCTATGAGGAATATCTGGGATGGACACCA 3196  
996 ysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGluLysArgGly 1012  
3197 CGAAGGCCAATAGTTTGGGAGTCACATGGCACAAA.....ATGAGGCAA 3240  
1013 LeuAlaAlaAsnProGlySerArgTyr.....AspAspLeuValSerAr 1027  
3241 ATGTGGATACAAAATGACCTTTATCTGGATGATGACCATTAGTAA 3290  
1027 g.....LeuLeuGluGlnGlyTyrTrpProGlyGluL 1038  
3291 CCAGCCTTCTTGAGAGCTCTGTAGGCCACTGCAGCAATTCCTGACGAA 3340  
1038 euLeuAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrSer 1054  
3341 GCACCACTCTGGGAGTTGAAGAATAGCAGTTTGAAGCAGCAGTTAA 3390  
1055 GluGluAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMe 1071  
3391 CAA.....GGAGCA.....TATAGCAT 3407  
1071 tValThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGln 1088  
3408 GGATACGCCAG.....TTGTGATGAGCTCATAGAACAATAGTACAG 3451  
1088 lnProGlu.....GluLeuArgAspLeuTyrSerLysHisLeuValAla 1102

```

1119 uLeuLysProSerGluArgAlaThrSerProValThrLeuSerProHisL 1136
1136 yHisValSerGlyPheSerSerLeuArgThrSerSerThrGlyAsp 1152
3577 .....GCTCCTCAGCTCACA.....GGGGA 3597
1153 AlaGlyGlySerArgArgProHisArgLys..... 1163
3598 ACAGGAGTGTGCCAAGAGTCGCATGCAAACTCAGCAAGCTGAC 3647
1164 .....ProThrIleLeuArgLysIleSerAlaAlaGlnG 1175
3648 ATTCAGCCGAAAGAGCTGTCTTCATGAGGCAA...AGCCANCTCCCT 3694
1175 InLeuSerAlaSerGluValVal...ThrHisLeuGlyGlnThrValAla 1190
3695 CAATTTCATTATAAACAATAATATCCAGGATGGAAATACAGTATAC 3744
1191 LeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluAlaIleGlyHI 1207
3745 ATTACAAAAGGACAGAGGAGTCATCAATATACTGTGTGACCTTATTACCC 3794
1207 sProArgProThrIleSerThrAlaArgAsnGlyGluGluValGlnPheS 1224
3795 CAGTGAGGCACATATACATGAGGACCAAGGATGGAACTTGTACAGCCCT 3844
1224 erAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaPro 1240
3845 CAGTAAATAATTTTGGATGGAACTGGGAGATACAGATACAGAACTCT 3894
1241 ValGluAlaAspValGlyPheThrCysAsnAlaThrAsnAlaLeuGL 1257
3895 ACAAGGAAAGAACCAAGCATATATGAATGTCGTAGCTAATCATCTGG 3944
1257 yThrAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
3945 TTCAGATGGGAAGTCTCTGTGTCGTATGCAGAGGCACCTGTCACT 3994
1274 yThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
3995 TGTCTGTGAAAGAAATATCACCACCAAGCAGCAGCACCACTGTCTGTT 4044
1291 AspIleGlySerThrIleLysThrValGlnGlyValAsnValThrIleAs 1307
4045 GTGGTGGAGGCATCTGGAGGAGCCCTTGGAGCAACCTGCAATCCG 4094
1307 nCysGlnValAlaIaGlyValProGluAlaGluValThrTrpPheArgAsnL 1324
4095 ATGCTCTGTAAGAGTGTCCCTCAGCCTAATAATACTTGGTTGAAGAG 4144
1324 ySerLysLeu...GlySerProHisHisLeuHisGluGlySerLeuLeu 1339
4145 GAGGATCTCTGAGTGGCAATGTTTCTGCTTTCATGGATCCCTGTTCG 4194
1340 LeuThrAsnValSerSerSerAspGlnGlyLeuThrSerCysArgAlaL 1356
4195 TTGCAGAAATGTTTCCCTTGAAATGAAGAACTAGCTGTGCTGATGCCAC 4244
1356 aAsnLeuHisGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAsp 1373
4245 CAATGCTCTTGGNAAGGAGTGGCACAATCTGTATTCCACTTGTCTGGAAC 4294
1373 roProGlnValProThrGlnLeuGluAspIleArgAla..... 1385

```

```

4295 GAAGATGGCCAGAGAGTAGAATCGTATTTCGCAAGGACATAAAAGTAC 4344
1386 LeuLeuAlaAlaThrGlyProAsnLeuProSerValLeuThrSerProLe 1402
4345 ATCTCCAGGCAACCAACACACTAGAACCAACAGC..... 4377
1402 uGlyThrGlnLeuValLeuAspProGlyAsnSerAlaLeuLeuGlyCysP 1419
4378 .....AATGACCCA..... 4386
1419 roIleLysGlyHisProValProAsnIleThrTrpPheHisGlyGlyGln 1435
4387 .....ACAGGAGAACCCCGCCT..... 4404
1436 ProIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGlyGI 1452
4404 ..... 4404
1452 nIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyGluPheSerC 1469
4404 ..... 4404
1469 ysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuVal 1485
4404 ..... 4404
1486 IleGlnAspTyrTrpTrpSerValAspArgLeuAlaThrCysSerAlaSe 1502
4405 ...CAGAGACCTTTTGGAGCCTGGTAACGTGTCACATTTGTTCTGCCAC 4451
1502 rCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLeu...A 1518
4452 CTGTGGTCAATTGGAGCCCGCAITCAGAGACCCCGGTGTGTGATGGCCA 4501
1518 snSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProAla 1534
4502 ATGGCAGAGAAAGTGTGAGGCCCTGTGTGATCACCCTCCAGAGCCACTG 4551
1535 ValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMetVa 1551
4552 GCTGGGTTTGAGCCCTGTATACATCCGGGACTGCCAGGAGGTGTTCCAC 4601
1551 lThrSerTrpSerAlaCysThrArgSerCysGlyGlyValGlnThrA 1568
4602 AAGTGTGTGTCACAGTCTCTGTGTCTGTGCGTGAGGATACACAGTC 4651
1568 rgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProVal 1584
4652 GGCAGGTGAGTGCAGCGGACAAAGCAATGCACTGTCCAGGTGGTG 4701
1585 SerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGlnAl 1601
4702 TCTCCAGAGCAGTGTGCCCT...AAGACCGGCCTCTGGGAAGAAAC 4748
1601 aCysAsnGlnGlnLeuCysValGluTrpAlaPheSerSerTrpGlyGlnC 1618
4749 ATGTTTGGTCATCCATGTGTTCAGTGG.....GAACGAGGAAACCGGT 4792
1618 ysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnVal 1634
4793 GTCTGGACGCTGCAATGGCCCTGCTGTGAGGATGCAGCAGCGTCACACA 4842
1635 PheCysGlnThrArgAspGlyIleThrLeuProSerGluGlnCysSerAl 1651
4843 GCTGTCAACACACACAGCTCT.....GACTCCAACCTGTGTATGA 4880
1651 aLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysServ 1668
4881 CAGAAAGACGCCCTTATAGAGGAGACTGCACATCAGGGGCTGTGTATG 4930
1668 alHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyAsn 1684
4931 TGTGTTGGCACACAGGCCCTTGGAAAGCCCTGTACAGCAGCCTGTGGCAG 4980

```

1685 TyrGlyPheGlnSerArgArgValGluCysValHisAlaArgThrAsnLys 1701  
 1701 sAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpG 1718  
 5028 ACCTGTGGCCAGAGACACTGTGTACAGAAAAGAAACCAATTTCTGGC 5077  
 1718 lnArgCysAsnIleThrProCysGlnAsnMetGluCysArgAspThrThr 1734  
 5078 GGACACTGT...CTGGGCCCTCTGTGTATAGAGACTGCACAGACAACT 5124  
 1735 ArgTyrCysGlnLysValHisGlnLeuLysLeuCysGlnLeuSerGlnPh 1751  
 5125 CACTACTGTATGTTGTAAACATCTTAATTTGTGTCTCTAGACGGCTA 5174  
 1751 eLysSerArgCysCysGlyThrCys 1759  
 5175 CAACCAAGGTGCTGCCAGTCATGT 5199

seq name: gb\_pat:AX224823

seq\_documentation\_block:  
 LOCUS AX224823 4854 bp DNA linear PAT 10-SEP-2001  
 DEFINITION Sequence 15 from Patent WO0161011.  
 ACCESSION AX224823  
 VERSION AX224823.1 GI:15554916

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 4854)  
 AUTHORS Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.  
 and Sands,A.T.  
 TITLE Novel human thrombospondin repeat proteins and polynucleotides  
 encoding the same

JOURNAL Patent: WO 0161011-A 15 23-AUG-2001;  
 Lexicon Genetics Incorporated (US)

FEATURES  
 source 1..4854  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 1331 a 1175 c 1276 g 1072 t  
 ORIGIN

alignment\_scores:  
 Quality: 3762.50 Length: 1733  
 Ratio: 3.194 Gaps: 31  
 Percent Similarity: 67.975 Percent Identity: 42.527

alignment\_block:  
 US-10-044-807-2 x AX224823 ..

Align seg 1/1 to: AX224823 from: 1 to: 4854

64 LysSerCysGluGlyArgAsnIleArgTyrArgThrCysSerAsnValas 80  
 94 AGGAATTTGTGAAGGCGACAAATTCGGTACAGAGATCGACAAATCATGA 143  
 80 pCysProProGluAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisA 97  
 144 CTGCCCTCCAGATGCAGAAAGATTTACAGAGCCAGCAGTGTCTCAGCTACA 193  
 97 snAspValHisHisGlyGlnPheTyrGluTrpLeuProValSerAsn 113  
 194 ATGATGTCCTCAGTATCAGGGCATTACTATGAAATGGCTTCCAGATATAT 243  
 114 AspProAspAsnProCysSerLeuLysCysGlnAlaLysGlyThrThrie 130  
 244 GATCCTGTGCCCGGTGTGCATCTCAAGTGTATGCACAGGACAAACTT 293

130 uValValGluLeuAlaProLysValLeuAspGlyThrArgCysTyrThrG 147  
 294 GGTGTGGAGCTGGCACTAAGGTACTGGATGGAACTCGTTGGCAACACGG 343  
 147 luSerLeuAspMetCysIleSerGlyLeuGlyCysGlnIleValGlyCysAsp 163  
 344 ACCTCTGGACATGTGTATCAGTGGCATCTGTACAGGAGTGGGTGGCAT 393  
 164 HisGlnLeuGlySerThrValLysGluAspAsnCysGlyValCysAsnGl 180  
 394 CGSACACTGGGAACAATGCCAGGAGGACAACTGTGGAGTCTGTGCCGG 443  
 180 yAspGlySerThrCysArgLeuValArgGlyGlnTyrLysSerGlnLeus 197  
 444 CGATGGCTCCACCTGCAGGCTGTACGGGGACAATCAAAAGTCACACGTT 493  
 197 erAlaThrLysSerAspAspThrValValAlaIleProTyrGlySerArg 213  
 494 CTCTGTGAANAAGAGAGAGAAATGTAATGCTGTCCTTTGGGAAGTCGA 543  
 214 HisIleArgLeuValLeuLysGlyProAspHisLeuTyrLeuGluThrLy 230  
 544 AGTGTGAGAATTACAGTGAAGGACCTGCCACCTCTTTATTAATCAAA 593  
 230 sThrLeuGlnGlyThrLysGlyGluAsnSerLeuSerSerThrGlyThrP 247  
 594 AACACTTCAAGGAAGCAAGGAGAACACAGCTTTAACAGCCCCCGCGTCT 643  
 247 heLeuValAspAsnSerSerValAspPheGlnLysPheProAspLysGlu 263  
 644 TTGTCGTAGANAACACACACAGCTGGAATTCAGAGGGCTCCGAGAGGCAA 693  
 264 IleLeuArgMetAlaGlyProLeuThrAlaAspPheIleValLysIleAr 280  
 694 ACTTTAAGATTCCAGGACCTCTGATGGCTGATTTTCATCTTCAGACCAG 743  
 280 gAsnSerGlySerAlaAspSerThrValGlnPheIlePheTyrGlnProI 297  
 744 GTACACTGCAGCCAAAGACAGCGTGGTTCAGTTCTTCTTTTACCAGCCCA 793  
 297 IleHisArgTirpArgGluThrAspPhePheProCysSerAlaThrCys 313  
 794 TCAGTCACTCAGTGGAGACAACTGACTTCTTCCCTGCACCTGTGACGTGT 843  
 314 GlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAs 330  
 844 GGAGGAGGTATATCAGCTCAATTCGTGAAIGTGTGGATATCCGCTTGA 893  
 330 nArgValValAlaAspGlnTyrCysHisTyrTyrProGluAsnIleLysp 347  
 894 GAGGAGTGTCTCTGACCATATTGTGACACTACTACCTGAAATGTAACAC 943  
 347 rOLysProLysLeuGlnCysAsnLeuAspProCysProAspSerAsp 363  
 944 CAANAACCAAACTGAAGGAATGCAGCATGGATCCCTGCCCATCAAGTGAT 993  
 364 GlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgThr 380  
 994 GGATTAAGAGATATATGCCCTATGACCACCTTCCAACTCTTCTCCGCTG 1043  
 380 pGluAlaThrProTirpThrAlaCysSerSerCysGlyGlyGlyIleG 397  
 1044 GGAACATAATCTTGGACATGCGATGTCGCTGTCCTGTGGAGGAGGATTC 1093  
 397 lnSerArgAlaValSerCysValGluAspIleGlnGlyHisValThr 413  
 1094 AGAGAGGGAGCTTGTGTGTGTAGAGAAATCCATGATGCATGAGAGATATTG 1143  
 414 SerValGluGluTirpLysCysMetTyrThrProLysMetProIleAlaGl 430  
 1144 CAGGTGGAAGATGGAAGTGCATGTACGCCACCAACCAAGGTATGCA 1193

430 nProCysAnllePheAspCysProLysTrpLeuAlaGlnGluTrpSerP 447  
1194 AACITGTAATCTGTTGATTTGCCCAAGTGGATTCGCATGGAGTGGCTC 1243  
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 453  
1244 ACTGCACAGTCACTGTGGCCGAGGGTTACGGTACCGGGTTGTCTGTGT 1293  
464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480  
1294 ATTAACCAACCGGAGAGACAGTGTGGGGCTGCATCCACACTGAAGTT 1343  
480 oHisIleLysGluGluCysIleValProThrProCysTyrLysProLysG 497  
1344 ACACATCAAGAAGAATGTGTCATTCCTCCATCCGTGCTTATAAACCAAAAG 1393  
497 LuLysLeuProValGluAlaLysLeuProTrpPheLysGlnAlaGlnGlu 513  
1394 AAAAAAGTCCAGTGAAGCAAAATTCCTTGGCTGAAACACAGCACAAAG 1443  
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530  
1444 CTAGAAGAGACCAAGTAATAGCAACAGAACCAAGTTCATTCAGAACCC 1493  
530 atTrpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547  
1494 CTGCTAGCCTGCAGTACCAAGTGTGGCCAGAGTGTGAGGTGCGCGAGG 1543  
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563  
1544 TGAAGTCCGTGTCTCTCACATTCACGCACAGCTGAGACTGAGCTGCC 1593  
564 IleAspGluCysGluGlyProLysProAlaSerGlnArg..... 576  
1594 GAGGAAGAGTGTGAAGGCCCAAGCTGCCCCAGCAAGCGCCTGCCTCCT 1643  
577 ....AlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsnP 592  
1644 GGAAGCATGTGATGAGAGCCCGGCTCCGAGAGCTAGACATCCCTCTCC 1693  
592 roAspGluThrAspGlyLeuPheGlyLeuGlnAspPheAspGluLeu 608  
1694 CT.....GAGGACAGTGAAGACGACT 1713  
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG1 625  
1714 TAGCACTGGAGTACGCTGGGTTACCCCTTGCCACAGCAACATGCTTGG 1763  
625 YGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGluP 642  
1764 AGGCCATCAAGAAGCCATAGCACTGTGCTTACATATCCAGACCCAGAGA 1813  
642 roAlaGluGluAsnLeuCysValThrSerArgArgProProGluLeuLeu 658  
1814 CAGTCAATCACAGCTGTGTGATATGTCACCGTCTCTCCAGCCATGAGC 1863  
659 LysSerCysAsnLeuAspProCysProAlaArgTrpGluIleGlyLysTr 675  
1864 CAGGGCTGTAAACACAGAGCCCTGCCCCAGGTGGCATGTGGGCTCTTG 1913  
675 pSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValP 692  
1914 GGGGCCCTGCTCAGCTACTGTGGAGTTGGAATTCAGACCCGAGATGTGT 1963  
692 heCysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAla 708  
1964 ACTGCTTGAC.....CCAGGGAGAGCCCTGCCCTGCCCT 1998  
709 AspGluLeuCysArgGlnProLysProSerThrValGlnAlaCysAsnAr 725  
1999 GAGGAG...TGCCGAGATGAAAAGGCCCATGCTTTACAGCATGCAATCA 2045  
725 gpHeAsnCysProProAlaTrpTyrProAlaGlnTrpGlnProCysSerA 742

2046 GTTGTGACTCCCTCCGCTGGCAGCATTTGAAGAATTGGCAGCAGTGTTCGA 2095  
742 rgThrCysGlyGlyGlyValGlnLysArgGluValLeuCysLysArg 758  
2096 GGACTTGTGGCGGGCAACTCAGAACAGAGTCACTGTGGCAGCTG 2145  
759 MetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlaSe 775  
2146 CTAAACGGATGGCAGCTTTTGAATCTCTCAGTAAGATTGTGCCAAGACC 2195  
775 rLysProAlaCysGlnGlnAlaCysLysLysAspCysProSerGluT 792  
2196 CAAGGCATGCTCACAAGTCTGTGCCAGCAGACATGCTCTCCACATT 2245  
792 rpLeuLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyThr 808  
2246 TAGCTGTGGAGACTGTCGAAGTCTTCTGTCACTGTGTGGTGGATC 2295  
809 GlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerTh 825  
2296 CAGAGAAGAAAGAGGTGTCTAAAGGTGGCAGCCAAAGAGTCGGCGCAT 2345  
825 rValValAsnSerThrLeuCysProProLeuProPheSerSerSerIleA 842  
2346 CCCCTCAGTCAGATGATGTCAGGGATCTACCAAGGTGCCCTTGTGTAA 2395  
842 rpProCysMetLeuAlaThrCysAlaArg...ProGlyArgProSerThr 857  
2396 GATCTTGCCAGATGCTCAGTGCAGTAAATCAATCAGAGATCAAGACA 2445  
858 Lys.....HisSerProHisIleAlaAlaArgLysValTyrIl 871  
2446 AAATCTGTGGACGAGGTCCGACATCTCAGTGTCCAGAGAGTCTACAT 2495  
871 eGlnThrArgArgGlnArgLysLeuHisPheValValGlyClyPheAlaT 888  
2496 TCAGACAAGGAAGAGAGCGGTATTAACTTGACCATTTGGTAGCAGAGCCT 2545  
888 yrLeuLeuProLysThrAlaValValLeuArgCysProAlaArgArgVal 904  
2546 ATTTGCTGCCCAACACATCCGTGATATTAAAGTCCCGCTGCCACGATTC 2595  
905 ArgLysProLeuIleThrTrpGluLysAspGlyGlnHisLeuIleSerSe 921  
2596 CAGAAATCTGTATCCCACTGGGAGAGGATGGCGGTTCGTCGAGAATC 2645  
921 rThrHisValThrValAlaProPheGlyTyrLeuLysIleHisArgLeuL 938  
2646 CAACGGCTGGCATCACCAAGTCAGGCTCACATAAAATCCACGGTCTTG 2695  
938 ysProSerAlaGlyValTyrThrCysSerAlaGlyProAlaArgGlu 954  
2696 CTGCCCCGACATCGGCGTGTACCGGTGCATTCGAGCTCTGCACAGGAA 2745  
955 HisPheValIleLysLeuIleGlyLysAsnArgLysLeuValAlaArgPr 971  
2746 ACAGTTGTCTCAAGTCAATTGTGTACTGACACACGGCTCATCGCAGCCC 2795  
971 oLeuSerProArgSerGluGluValLeuAlaGlyArgLysGlyGlyP 988  
2796 A.....GCCCTCAGGGAGCGCTATCA 2815  
988 roLysGluAlaLeuGlnThrHisLysHisGluAsnGlyIlePheSerAsn 1004  
2816 GGGAATATCTGGATGGACCGACAGCGCAACCCAAATGTTTGGAGTCA 2865  
1005 GlySerLysAlaGulysArgGlyLeuAlaAlaAsnProGlySerArgTy 1021  
2866 TGGCAGAAA.....ATGAGCAAAATGTGGTAATACAAAATGACCTTTA 2909  
1021 r.....AspAspLeuValSerArg.....LeuLeu 1030

[illegible]





1268 GCTGAATGTGGTATATCCCTTGAAGAGGGTAGTTCCTGACCATATTG 1317  
 338 sHisTyrTrpProGluAsnIleLysProLysLeuGluGluCysA 355  
 1318 TCACTACTAGCCTGAAATGTAACCAACCAAACTGAAGGAATGCA 1367  
 355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371  
 1368 GCATGGATCCCTGCCATCAAGTATGATGATTTAAAGAGATAATGCCCTAT 1417  
 372 AspLeuTyrHisProLeuProArg..TrpGluAlaThrProTyrThrAla 387  
 1418 GACCACCTCCAACTCTCCTCGAGCTGGGAACATAATCCTGGACTGCA 1467  
 388 CysSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerCysVa 404  
 1468 TGTTCCTGTCTGTGGAGGAGGATTACAGAGCGGAGCTTGTGTGTGT 1517  
 404 lGluGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysM 421  
 1518 AGAGAAATCCATGATGGAGAGATATTCCAGGTGGAGAAATGGAAGTCA 1567  
 421 eTyrThrProLysMetProIleGlnProCysAsnIlePheAspCys 437  
 1568 TGTAGCGACCCAAACCCAAAGTTATGCAAACTGTGAATCTGTTGATTGC 1617  
 438 ProLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyG 454  
 1618 CCCAAGTGGATTGGCATGGAGTGGTCTCAGTGCACGACTGTGGCGCG 1667  
 454 nGlyLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHist 471  
 1668 AGGTTACGTTACCGGGTGTCTGTGTATTAAACCACCGCGAGAGCATG 1717  
 471 hrGlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIle 487  
 1718 TTGGGGCTGCAATCCAACTGAAGTTACATCAAAAGAAATGTGTC 1767  
 488 ValProThrProCysTyrLysProLysGluLysLeuProValGluAlaL 504  
 1768 ATTCCTCATCCGTTGTTATAACCAAAAGAAAAGTCCAGTGGAGCAAA 1817  
 504 sLeuProTrpPheLysGlnAlaGlnGluLeuGluGlyAlaAlaValS 521  
 1818 ATTGCTTTGGCTGAACACAGCACAGAACTAGAACAGCACCAATAGCAA 1867  
 521 erGluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThr 537  
 1868 CAGAAAGAACCAACGTTTCATTCAGAAACCTGTGTACCCCTGCAGTACCAG 1917  
 538 CysGlyValGlyThrGlnValArgIleValArgCysGlnValIleLeuSe 554  
 1918 TTGTGGCCAGGTGTGCAGTCCGCGAGGTGAAGTCCGCTGTCTCCTCAC 1967  
 554 rPheSerGlnSerValAlaAspLeuProIleAspLeuCysGluGlyProL 571  
 1968 ATTCAGCGACAGTGAAGACTGAGTGTCCCGCAGGAAGAGTGTGAAGGCCCA 2017  
 571 ysProAlaSerGlnArg.....AlaCysTyrAlaGlyPro 582  
 2018 AGCTGCCACCGAACCGCCCTGCCTCTGGAAGCATGTGATGAGAGCCCG 2067  
 583 CysSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPh 599  
 2068 GCCTCCCGAGAGTAGACATCCCTCTCCCT..... 2097  
 599 eGlyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyP 616  
 2098 .....GAGACAGTGAAGACGACTTACGACTGGGAGTACGCTGGGT 2137  
 616 heThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValVal 632  
 ||||| ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||

2138 TCACCCCTTGCACAGCAACATGCTTGGAGGCCATCAGNAGCCATAGCA 2187  
 633 SerCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysVa 649  
 2188 GTGTGCTTACATATCCAGACCCAGACAGTCAATGACAGCTTGTGTGA 2237  
 649 lThrSerArgProProGlnLeuLeuLysSerCysAsnLeuAspProC 666  
 2238 TATGTCCACCGCTCTCCAGCCATCAGCCAGGCGCTGTAACACAGAGCCT 2287  
 666 ys.Pro..... 667  
 2288 GTCCCCCAGGAGAGAGCCAGCAGCTTGTAGNAGCATGCCGGTTACATA 2337  
 667 ..... 667  
 2338 ATGTCTCTGTAGTCTGAGGAGAGCGCTTCTTCTTAACAGGATTCAACAC 2387  
 667 ..... 667  
 2388 TGCTAGGAAGAAAGAGGAGAAAGAGGCAATAGTGTGTGTTTCTGT 2437  
 667 ..... 667  
 2438 ACCAGCTTGTACCTATTCTTGTATATAAAAAACAATCTTTATTGAGTT 2487  
 667 ..... 667  
 2488 CATTTCTGTGAATAAGAAATTTGCCCATTTCTTAATAAAAAACAGCT 2537  
 668 .....AlaArgTrpGluIleGlyLysTrpSerPro 677  
 2538 CCATCTCCAAAAAATAAATGGCATGTGGGCTTTGGGGGCC 2587  
 678 CysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValPheCysSe 694  
 2588 TGCTCAGCTACCTGTGGAGTTGGAATTCAGACCGAGAGTGTACTGCCT 2637  
 694 rHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAlaAspGluL 711  
 2638 GCAC.....CCAGGGGAGACCCCTGCCCTCTCTGAGGAG. 2671  
 711 euCysArgGlnProLysProSerThrValGlnAlaCysAsnArgPheAsn 727  
 2672 ..TGCGAGATGAAAGCCCATGCTTTACAGCATGCAATCAGTTTGAC 2719  
 728 CysProAlaTrpTyrProAlaGlnTrpGlnProCysSerArgThrCy 744  
 2720 TGCCCTCTCTGGTGGCAGCATTTGAAGAATGGCAGCAGTGTCCAGGACTTG 2769  
 744 sGlyGlyValGlnLysArgGluValLeuCysLysGlnArgMetAlaA 761  
 2770 TGCGGGGGAACCTCAGACAGAGAGTCACTGTCCGAGCTGCTAACGG 2819  
 761 spGlySerPheLeuGluLeuProGluThrPheCysSerAlaSerLysPro 777  
 2820 ATGGCAGCTTTTGAATCTCTCAGATCAATTTGCCAAGAGACCCAGGCA 2869  
 778 AlaCysGlnGlnAlaCysIlyLysAspCysProSerLysTrpLeuLe 794  
 2870 TCTCTCACAAAGTCTGTGCCAGACAGACTGTCTCCACATTAGCTGT 2919  
 794 uSerAspTrpThr..... 798  
 2920 GGGAGACTGTGTCGAGGAGCATTAATGCAAGAGAGCAATGGAGCAGGAT 2969  
 798 ..... 798  
 2970 CTACACAAATCTTAAGAAAGCAAGCATGACTCAAGGATTTCTCTTCAT 3019  
 798 ..... 798  
 3020 CCTGTGTTCTTCATGTAGAGAGACAGAGAGGAGGAGTCTCAGAGAAATCTG 3059



[illegible]



```
234 lYrThySgLyGluAnSerLeuSerSerThrGlyThrPheLeuValAsp 250
|||||
751 GGACTAAAGGTGAAACAGTCTCAGCTCCACAGAACTTTCTTGTGGAC 800
|||||
251 AsnSerSerValAspPheGlnLysPheProAspLysGluLeuLeuArgMe 267
|||||
801 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATACTGAGAAT 850
|||||
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
851 GGCTGGACCACTCACACAGATTCATGTCAAGATTCGTAATCGGGCT 900
|||||
284 erAlaAspSerThrValGlnPheIlePheTyrglnProIleIleHisArg 300
|||||
901 CCGCTGACAGTACAGTCCAGTTCATCTTCATCAACCCATCATCCACGA 950
|||||
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyTy 317
|||||
951 TGGAGGAGAGCGGATTTCTTTCTTGTCTCAGCAACCTGTGGAGAGGTTA 1000
|||||
317 rGlnLeuThrSerAlaGluCysTyrglnLeuArgSerAsnArgValVala 334
|||||
1001 TCACCTGACATCGCTCAGTCTACGATCTCAGGAGCAACCGTGTGTTG 1050
|||||
334 laAspGlnTyrCysHisTyrglnProGluAsnIleLysProLysProLys 350
|||||
1051 CTGACCAATAGTCTCACTATTACCCAGAGAAATCAAAACCCAAACCCAA 1100
|||||
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367
|||||
1101 CTTCAGGAGTGCACTTGGATTCCTGTGTCAGCCAGTGACGGATACAGCA 1150
|||||
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTrpGluAlaThrP 384
|||||
1151 GATCATGCCATTACCTTACCATCCCTCTCTCGTGGGAGGCCACCC 1200
|||||
384 roTrpThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
|||||
1201 CATGGACCGCGTGTCTCTCCCTCGTGTGGGGGGGCATCCAGAGCGGGCA 1250
|||||
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417
|||||
1251 GTTTCCTGTGTGGAGGAGGACATCCAGGGCATGTCACTTCAGTGGAAAG 1300
|||||
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1301 GTGGAATGTCATGTACACCCCTAAGATGCCCATCGCGCAGCCCTGCAACA 1350
|||||
434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
|||||
1351 TTTTGTGTCGCCCTAAATGGCTGGCACAGGAGTGTCTCCGTGTCACAGTG 1400
|||||
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1401 ACGTGTGGCCAGGCGCTCAGATACCGTGTGTCTCTCATCGACCATCG 1450
|||||
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1451 AGGAATGCACACAGGAGCGCTAGGCCAAAAACAAAGCCCAATATAAAG 1500
|||||
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1501 AGGAATGATCGTACCCATCCCTGCTATAACCCAAAGAGAAATCTTCCA 1550
|||||
501 ValGluAlaLysLeuProTrpPheLysGlnAlaGlnGluLeuGluGluG 517
|||||
1551 GTGAGGCGCACTTGCATGTGTCAACAGCTCAAGAGCTAGAGAAGG 1600
|||||
517 yAlaAlaValSerGluGluProSerPheIleProGluAlaTrp..... 531
|||||
1601 ACCTGCTGTGTAGAGGAGCCCTCGTAAGTTGTAAGACACACAGACTGTTC 1650
|||||
532 .....SerAlaCysThrValThrCys 538
```

```
|||||
1651 TATATTGAAACTTTTGTTTAAAGAAACAGTGTCTCACTGGTGTAGCT 1700
|||||
539 GlyValGlyThrGlnValArgIleVal 547
|||||
1701 TTCATGGTCTGAACTAAGTGTATC 1727
|||||
seq_name: gb_pat:AX224819
seq_documentation_block:
LOCUS AX224819 2538 bp DNA linear PAT 10-SEP-2001
DEFINITION Sequence 11 from Patent WO0161011.
ACCESSION AX224819
VERSION AX224819.1 GI:15554914
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2538)
AUTHORS Donoho,G., Scovillie,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
TITLE Novel human thrombospondin repeat proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0161011-A 11 23-AUG-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source 1..2538
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 670 a 627 c 678 g 563 t
ORIGIN
alignment_scores:
Quality: 2809.50 Length: 783
Ratio: 4.302 Gaps: 4
Percent Similarity: 83.397 Percent Identity: 60.920
alignment_block:
US-10-044-807-2 x AX224819 ..
Align seg 1/1 to: AX224819 from: 1 to: 2538
22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
|||||
190 ACCTCAAGAAACACCTCGTTCAGATGAAGACAAAGATGGCACTGGGATGC 239
|||||
38 aTrpGlyProTrpSerGluCysSerArgThrCysGlyGlyAlaSerT 55
|||||
240 TTGGGGGCACTGGAGTGACTGCTCCCGACCTGTGGGGGAGGAGGATCAT 289
|||||
55 yTrSerLeuArgArgCysLeuSerSerLysSerCysGluGlyArgAsnIle 71
|||||
290 ATTCCTCGGAGATGTTGACTGGAGGAAATGTGAAGGGCAGAACATT 339
|||||
72 ArgTyrArgThrCysSerAsnValAspCysProGluAlaGlyAspPh 88
|||||
340 CGGTACAAGACATGCGCAATCATGACTGCCCTCCAGATGCAGAGATT 389
|||||
88 eArgAlaGlnGlnCysSerAlaHisAsnAspValLysHisGlyGlnP 105
|||||
390 CAGAGCCCGACGAGTGTCTCAGCTCAATGATGTCCAGTATCAGGGGCAT 439
|||||
105 heTyrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
|||||
440 ACTATGAATGGCTTCCACGATATATGATCTGCTGCCCGGTGTGCATC 489
|||||
122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138
|||||
490 AAGTGTGTCATGCACAGGACAAACACTTGGTGGAGCTGGCACCTAAGGT 539
|||||
138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
```

```

|||||
540 ACTGATGGAACCTCGTTGCAACACGACCTCTGGACATGTGTATCAGTG 589
|||||
155 lYLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrVallys 171
|||||
590 GCATCTGTGACGACAGTGGCTGCGATCGCACTGGGAAGCAATGCCAAG 639
|||||
172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgIleuVa 188
|||||
640 GAGGACAACTGTGGAGTCTGTCCGCGCATGGCTCCACCTGCAGGCTGT 689
|||||
188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspThrV 205
|||||
690 ACGGGACAACTCAAGTCAACAGCTTCTCCCTGAAAAAAGAGAAATG 739
|||||
205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221
|||||
740 TAATGCTGTTCTTGGGAAGTCGAAGTGCAGAAATACAGTGAAGGA 789
|||||
222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyG 238
|||||
790 CCTGCCCACTCTTATTGAAATCAAAAACACTTCAAGGAAGCAAGGAGA 839
|||||
238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerValA 255
|||||
840 ACACAGCTTTAACAGCCCGCGCTTTGCTGTAAGAAACACACAGTGG 889
|||||
255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271
|||||
890 AATTTCAGAGGGGTCCGAGAGGCAAACTTTAAGATCCAGGACCTCTG 939
|||||
272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288
|||||
940 ATGCTGATTTCACTTCAAGACAGGTACACTGCAGCAAGACAGCGT 989
|||||
288 rValGlnPheIlePheTyrGlnProIleIleHisArgTrpArgGluThra 305
|||||
990 GGTTCAGTCTCTTTTACCAGCCCATCAGTCATCAGTGGAGACAAACTG 1039
|||||
305 spPhePheProCysSerAlaThrCysGlyGlyTyrGlnLeuThrSer 321
|||||
1040 ACTTCTTCCCTGCACTGTGACGTGTGGAGAGGTATATCAGTCAATCT 1089
|||||
322 AlaGlyCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
|||||
1090 GCTGAATGTGTGATATCCGCTTGAAGAGGGTAGTTCTTGACCATTTG 1139
|||||
338 sHisTyrTyrProGluAsnIleLysProLysProLysLeuGlnLucysA 355
|||||
1140 TCACCTACTACCTGAAATGTAAACCAAAACCAAACTGAAGAAATGCA 1189
|||||
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
|||||
1190 GCATGGATCCCTGCCCATCAAGTATGATGGATTTAAAGAGATAATGCCCTAT 1239
|||||
372 AspLeuTyrHisProLeuProArgTrpGluAlaThrProThrAlaCy 388
|||||
1240 GACCACCTCAACCTCTCTCTCGCTGGAAACATAATCCTTGGACTGCATG 1289
|||||
388 sSerSerSerCysGlyGlyIleGlnSerArgAlaValSerCysValG 405
|||||
1290 TTCCGTGCTGCTGGAGAGGATTCAGACACGAGCTTTGTGTGTGTAG 1339
|||||
405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421
|||||
1340 AGGAATCCATGTCATGGAGAGATATTGACAGTGGGAAGAAATGGAATG 1389
|||||
422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438
|||||
1390 TAGCACCCCAACCCCAAGGTTTGAACAACTGTAATCTGTTGATGCC 1439
|||||
438 oLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455
|||||

```

```

1440 CAAGTGGATTGCCATGGAGTGGTCTTCAGTGCACAGTGACTTGTGGCGGAG 1489
|||||
455 lYLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
|||||
1490 GGTACGGTACCGGGTGTCTGTGTATTAAACACCGCGAGAGCATGTT 1539
|||||
472 GlyGlyCysSerProLysThrLysProHisIleLysGluLucysIleVa 488
|||||
1540 GGGGCTGCATCCCACTGAAGTTACATCAAAAGAAGAAATGTGTCTCAT 1589
|||||
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
|||||
1590 TCCCATCCCGTGTATATAAACCAAAAGAAAAAGTCCAGTGGAAAGCAAAAT 1639
|||||
505 euProTrpPheLysGlnAlaGlnLeuGluGluGluGlyAlaLaValSer 521
|||||
1640 TGCCCTGGCTGAAACAAGCACAGAACTAGAAGAGACCAAGATAGCAACA 1689
|||||
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538
|||||
1690 GAAGAACCACCTTCATTCACAACCCCTGTCAGCTGCAGTACCACGTG 1739
|||||
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
|||||
1740 TGGGCGAGGTGTCAGTCCGCGAGGTGAAGTCCGCTGTGCTCCTCACAT 1789
|||||
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys 571
|||||
1790 TCACCGAGACTGAGACTGAGCTGCCCGAGGAGAGGTGTGAAGGCCCAAG 1839
|||||
572 ProAlaSerGlnArg. ....AlaCysTyrAlaGlyProCy 583
|||||
1840 CTGCCACCGAAGCGCCCTCCCTCTCGGAGCATGTGATGAGAGCCCGCG 1889
|||||
583 sSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheG 600
|||||
1890 CTCCCGAGAGCTAGACATCCCTCTCCCT. .... 1917
|||||
600 lYGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPhe 616
|||||
1918 .....GAGGACAGTGAGAGACTTACGACTGGGAGTACGCTGGGTTC 1959
|||||
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValSe 633
|||||
1960 ACCCTTCACAGCAACATGCTTGGAGGCCATCAAGAACCTATAGCAGT 2009
|||||
633 rCysLeuAsnLysGlnThrArgGluProAlaGluAsnLeuCysValT 650
|||||
2010 GTGCTTACATATCCAGACCCAGACAGTCAATGACAGCTTGTGTGATA 2059
|||||
650 hrSerArgArgProGlnLeuLysSerCysAsnLeuAspProCys 666
|||||
2060 TGTTCACCGTCTCCAGCATGAGCCAGGCCCTGTAAACAGAGCCCTGT 2109
|||||
667 ProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysGl 683
|||||
2110 CCCCCCAGGTGGCATGTGGGCTCTTGGGGCCCTGCTCAGCTACCTGTGG 2159
|||||
683 yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArg 700
|||||
2160 AGTTGGAATTCAGACCCGAGATGTGTACTGCTGCAC. .... 2196
|||||
700 luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716
|||||
2197 ..CCAGGGGAGACCCCTGCCCTCCTGAGGAG...TGCCGAGATGAAGAG 2241
|||||
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733
|||||
2242 CCCCATGCTTACAAAGCATGCAATCAGTTGACTGCCCTCCTGTGCTGGCA 2291
|||||
733 rProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnL 750
|||||
2292 CATTGAAGAATGGCAGCAGTGTCCAGGACTTGTGGCGGGGGAACCTCAGA 2341
|||||

```

Percent Similarity: 83.961 Percent Identity: 62.483

alignment\_block:

US-10-044-807-2 x AF237652 ..

Align seg 1/1 to: AF237652 from: 1 to: 2523

22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38

414 ACCTCAAGAAACACTCGTTTCAGATGAGACAAAGATGCGCACTGGATGC 463

38 aTrpGlyProTrpSerGluCysSerArgThrCysGlyGlyAlaSerT 55

464 TTGGGGCCACTGGAGTGACTGCTCCCGACCTGTGGGGAGGAGCATCAT 513

55 yrSerLeuArgArgCysLeuSerSerCysGluGlyArgAsnIle 71

514 ATTCTCTCGGAGAGATTGTGAGGAAATTTGAGGGGAGGAGCAACATT 563

72 ArgTyrArgThrCysSerAsnValAspCysProProGluAlaGlyAspPh 88

564 CGGTACAAGACATGACGCAATCATGACTGCCCTCCAGATGCAAGAGATT 613

88 eaGAlaGlnGlnCysSerAlaHisAsnAspValLysHisHisGlyGlnP 105

614 CAGAGCCAGCAGTGTCTCAGCTACATGATGCTCAGTATCAGGGGCAATT 663

105 heTyrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121

664 ACTATGATGATGGTCTCCAGATATATATGATGCTGCTCCCGGTGCACATC 713

122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138

714 AAGTGTCTATGCACAAAGCAAACTTGGTGTGGAGCTGGCACTAAAGT 763

138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155

764 ACTGGATGGAACCTCGTTGCAACAGGACCTCTGGACATGTGTATCAGTG 813

155 lLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLys 171

814 GCATCTGTCCAGCAGTGGGTGCGATCGCAACTGGGAAGCAATGCAAG 863

172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188

864 GAGGACAACTGTGGAGTCTGCTCCCGGATGCTCCACCTGCGAGGCTGT 913

188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspAspThrV 205

914 ACGGGGACAACTAAAGTCAACAGTTCTCTGTAAGAAAGAGAGAAATG 963

205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLysGly 221

964 TAATTGCTGCTCTTGGGAAGTGAAGTGTGAGAATTACAGTGAAGGA 1013

222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGly 238

1014 CTGTGCTCACTCTTAAATGAAACAAACACTTCAAGGAAGCAAGAGGA 1063

238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerVal 255

1064 ACACAGCTTTAAAGCCCGGCGCTTTTGTGCTAGAAAACACACACAGTGG 1113

255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271

1114 AATTTTCAGAGGGCTCCGAGAGGCAAACTTTTAAAGATTCCAGGACCTCTG 1163

272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288

1164 ATGCTGATTTCACTTCAACAGCAGGATACACTGCAGCCCAAGACAGCGT 1213

288 rValGlnPheIlePheTyrGlnProIleIleHisArgTrpArgGluThra 305

750 ysArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGlu 766

2342 ACAGAAGACACCTGTCGGCAGCTGCTAAGGATGCGACTTTTTCAT 2391

767 LeuProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCy 783

2392 CTCTCAGATGAATTTGTCAGGACCAAGGATGCTCTCAGAGTCTG 2441

783 sLysLysAspAspCysProSerGluTyrLeuLeuSerAspTrpThrGlu 799

2442 TGCCAGACAGACAGTCTCTCAGATTAAGTGTGGGAGACTGGTCGAG 2490

seq\_name: gb\_pr:AF237652

seq documentation\_block:

LOCUS AF237652 2523 bp mRNA linear PRI 02-MAR-2001

DEFINITION Homo sapiens a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3 (ADAMTSL3) mRNA, partial cds.

ACCESSION AF237652

VERSION AF237652.1 GI:13183077

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2523)

AUTHORS Hirohata, S., Anand-Apte, P., Seldin, M. and Apte, S.

TITLE Function, a member of a new family with similarities to ADAM-TS proteases, is a component of extracellular matrix of skeletal muscle

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2523)

AUTHORS Anand-Apte, P. and Apte, S.

TITLE Direct Submission

JOURNAL Biomedical Engineering, Cleveland Clinic Foundation, 9500 Euclid, Cleveland, OH 44195, USA

FEATURES

source

1..2523

/organism="Homo sapiens"

/db\_xref="taxon:9606"

1..>2523

/gene="ADAMTSL3"

/note="ADAMTSL3"

/note="ADAM-TSL3; similar to GenBank Accession Number AB033059; extracellular matrix protein"

/codon\_start=1

/product="a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 3"

/protein\_id="AAK15041.1"

/db\_xref="GI:13183078"

/translation="MASWTSPWVLIGVMFHSPLPOTTAEKSPGAFYLPFAISPOG SFLEDTGEGFLTYRYDDQTSRNTSRDEKDGNDWGDWSDCSRTCGGASYSLRRC LIGRNEGGQIRKTKNSHDCPDPAEDFAQQCSAYNDVOYQGHVYEWLPYNDPAAP CALKHAQGNLVEALPDKDTRCNDSLDMCISGICQAVGCDRLGDSNAKDNCG VCAAGDSFRLVGGSKSHVSPKREENVIAVPLGSRVIRTKVPVHLFIESKTLQG SKGEHSFSPGVVENVITVEFGSRQTFKIPPLMADFIKTRYTAADKSVQVF FYQPISHOWROTDFPCTVTTCGGYQINSAECVDIRLKVVPDHYCHYPENVKPKP LKCSMDPSSDGFKEIMPDHFOPLRWEHPWTACSVCGGIGIORSFVFCVEESM HGILQVEHKMKYAPKPKVMOTCNLFDCPKWIAEWSOCTVTCGRGLRYRVLCINH RGHVGCNQLKHLKEECVFIPICYKPKRSPVEAKLPWLKQALEETRTATEEP TFIPEWMSACSTTCGPGVQREVRCRVLLFTQTLEPPEECBPKLPFPCLLLEA CDSPASRELDIPLEPSETYDYWAGFTPCATCLGHQEAIAVCLHITQOTQVND SLGDMVRRHPAMSOACNETPCPRVHWGSGPCSATCGVGIQTRDNYCLHPGPAPP EECRDKPHALQACNDFCAPPWHIE"

BASE COUNT 635 a 652 c 692 g 544 t

ORIGIN

alignment\_scores:

Quality: 2625.50

Ratio: 4.361

Length: 717

Gaps: 4



```

1214 GGTTCAGTTCTTCTTTTACAGCCCAATCAGTCATCAGTGGAGACAAACTG 1263
305 spPheProCysSerAlaThrCysGlyGlyTyrGlnLeuThrSer 321
1264 ACTTCTTTCCTCCCTGCACTGACGTGTGGAGGAGTTATCAGCTCAATTCT 1313
322 AlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
1314 GCTGAATGTGGATATCCGCTGAAGAGGGTAGTTCCTGACCATATATG 1363
338 shiTyrtyrProGluAsnIleLysProLysProLysLeuGlnGluCysA 355
1364 TCACTACTACCTGAAATGTAACCAAAACCAAACTGAAGGAATGCA 1413
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
1414 GCATGATCCCTGCCATCAAGTGATGGATTAAAGAGATAATGCCCTAT 1463
372 AspLeuTyrHisProLeuProArgTrpGluAlaThrProTyrThrAlaCy 388
1464 GACCACCTCCCAACCTTCTCTCGCTGGGAACATAATCCTTGGACTGCATG 1513
388 sSerSerCysGlyGlyGlyLeuGlnSerArgAlaValSerCysValG 405
1514 TTCCGTGTCTGTGGAGAGGATTCAGAGCGGAGCTTGTGTGTGTAG 1563
405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421
1564 AGGAATCATCATCATGAGAGATATTGCAGGTGGAAGAATGGAAGTGCATG 1613
422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438
1614 TAGCCACCAACCAACCAAGTTATGCAACTTGTAACTCTGTGATTGGCC 1663
438 olYstTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455
1664 CAAGTGGATTGCATGGAGTGGTCTCAGTGCACAGTGACTTGTGGCCGAG 1713
455 lyLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
1714 GGTACGGTACCGGGTGTCTGTGTGTATTAACCAACCGGAGAGCATGTT 1763
472 GlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIleVa 488
1764 GGGGCTGCAATCCACAACACTGAAGTACATCAAGAAGAATGTGTGCAT 1813
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
1814 TCCATCCCGTGTATATAACCAAAAGAAAAGTCCAGTGAAGCAAAAT 1863
505 euProTrpPheLysGlnAlaGlnGluLeuGluGlyAlaAlaValSer 521
1864 TGCCTTGGCTGAAACAAGCAAGCAAGTAAAGAGAGACCAAGTAAAGCA 1913
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538
1914 GAAGAACAACAGTTTCATCCAGAACCTGGTCACCCCTGCAGTACCACGTG 1963
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
1964 TGGGCCGGGTGTGCAGGTCCGCGAGGTGAAGTGCCTGTCTCCATCAT 2013
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys 571
2014 TCAGCCAGACTGACACTGAGTCCGCCGAGGAAGAGTGTGAAGGCCCAAG 2063
572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProCy 583
2064 CTGCCCCACCAACCGCCCTGCTCTCCGGAAGCATGTGATGAGGCCCGGC 2113
583 sSerGlyGluIlePProGluPheAsnProAspGluThrAspGlyLeuPheG 600
2114 CTCGCCAGAGCTAGACATCCCTCTCCCT..... 2141

```

```

600 lyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPhe 616
2142 .....GAGGACAGTGAGACGACTTACGACTGGGAGTACGGTGGGTTTC 2183
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValSe 633
2184 ACCCTTGCACAGCAACATCTTGGGAGGCCATCAGAAGCCATAGCAGT 2233
633 rCysLeuAsnLysGlnThrArgGluProAlaGluAsnLeuCysValT 650
2234 GTGCTTACATATCCAGACCAGCAGACAGTCAATGACAGCTTGTGCATA 2283
650 hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys 666
2284 TGGTCCACCGCTCTCCAGCCATGAGCCAGGCTGTAAACAGAGCCCTGT 2333
667 ProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysG 683
2334 CCCCCAGGTGGCATGTGGCTCTTGGGGCCCTGCTACGCTACCTGTGG 2383
683 yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArg 700
2384 AGTTGAATTCAGACCCGAGATGTGTACTGCCTGCAC..... 2420
700 luMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716
2421 ..CCAGGGAGACCCCTGCCCTCTCTGAGGAG...TGCCGAGATGAAAG 2465
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733
2466 CCCCATGCTTTACAGCATGCAATCAGTTTGAATGCCCTCCCTGGCGCA 2515
733 r 733
2516 c 2516

```

seq\_name: gb\_pat:AX224821

seq\_documentation\_block:

LOCUS AX224821 2316 bp DNA linear PAT 10-SEP-2001  
 DEFINITION Sequence 13 from Patent WO0161011.  
 ACCESSION AX224821  
 VERSION AX224821.1 GI:15554915  
 KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2316)  
 AUTHORS Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.  
 and Sands,A.T.

TITLE Novel human thrombospondin repeat proteins and polynucleotides  
 encoding the same

JOURNAL Patent: WO 0161011-A 13 23-AUG-2001;  
 Lexicon Genetics Incorporated (US)

FEATURES  
 Location/Qualifiers

source

1..2316

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 617 a 575 c 616 g 508 t

ORIGIN

alignment\_scores:

Quality: 2615.50

Ratio: 4.267

Percent Similarity: 82.726

Percent Identity: 60.054

alignment\_block:

US-10-044-807-2 x AX224821 ..

Align seg 1/1 to: AX224821 from: 1 to: 2316

64	LysSerCysGluGlyArgAsnIleArgTyrArgThrCysSerAsnValAs	80
94	AGCAATTGTGAAGGGCAGAACATTCGGTACAGACATGACGCAATCATGA	143
80	pCysProProGluAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisA	97
144	CTGCGCCTCAGATGCAGAGATTTCAGAGCCGACGAGTGTCTAGCCCTACA	193
97	snAspValLysHisHisGlyGlnPheTyrGluTrpLeuProValSerAsn	113
194	ATGAUATCCAGATACAGGGCAATTACTATGAATGGCTTCCACGATATAAT	243
114	AspProAspAsnProCysSerLeuLysCysGlnAlaLysGlyThrThrLe	130
244	GATCTCTGTCGCCGCTGTGCATCAAGTGTCAATGCACAGAGCAAAACTT	293
130	uValValGluLeuAlaProLysValLeuAspGlyThrArgCysTyrThrG	147
294	GGTGTGGAGCTGGCACTTAAGTACTGGATGGAATCTGTGCAACACGG	343
147	IuSerLeuAspMetCysIleSerGlyLeuCysGlnIleValGlyCysAsp	163
344	ACTCTTGGACATGTCTATCAGTGGCATCTGTACGGCAGTGGGTGGGAT	393
164	HisGlnLeuGlySerThrValLysGluAspAsnCysGlyValCysAsnGl	180
394	CGCAACTGGGAAGCAATGCCAGGAGGACAACTGTGGAGTCTGTGGCGG	443
180	YAspGlySerThrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuS	197
444	CGATGCTCCACCTGCAGGCTGTGTACGGGACAAATCAAAATGCACACGTTT	493
197	erAlaThrLysSerAspAspThrValValAlaIleProTyrGlySerArg	213
494	CTCCTCGAAAAGAGAAAGAAATGTAATGCTGTCTCTTTGGGAAGTCGA	543
214	HisIleArgLeuValLeuLysGlyProAspHisLeuTyrLeuGluThrLy	230
544	AGTGTGAGAAATACAGTCAAGGACCTGCCACCTCTTATTGAATCAAA	593
230	sThrLeuGlnGlyThrLysGlyGluAsnSerLeuSerSerThrGlyThrP	247
594	AACACTTCAAGGAAGCAAAAGGAGAACACAGCTTTAAACGCCCGCGGTCT	643
247	heLeuValAspAsnSerSerValAspPheGlnLysPheProAspLysGlu	263
644	TTGTGCTGAGAAACACAAACAGTGGAAATTTACAGGGGCTCCGAGGGCAA	693
264	IleLeuArgMetAlaGlyProLeuThrAlaAspPheIleValLysIleAr	280
694	ACTTTTAAGATTCAGGACCTCTGATGCTGATTTCATCTTCAAGACCAG	743
280	gAsnSerGlySerAlaAspSerThrValGlnPheIlePheTyrGlnProI	297
744	GTACACTGCACCCAAAGACAGCGGTGTACAGTCTCTTTTACCAGCCCA	793
297	leIleHisArgTrpArgGluThrAspPhePheProCysSerAlaThrCys	313
794	TCAGTCAATCAGTGGAGACAACTCACTTCTTCCCTGCACGTGTGACGTGT	843
314	GlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAs	330
844	GGAGGAGGTATACGTCAAATCTCTGAATGTGGGATATCCGCTTGAA	893
330	nArgValValAlaAspGlnTyrCysHisTyrTyrProGluAsnIleLysP	347
894	CAGGGTAGTTCCTGACCAATTATTGTCTACTACTACCTGAAATGTAAAC	943
347	roLysProLysLeuGlnGluCysAsnLeuAspProCysProLysSerAsp	363
944	CAAAACCAAACTGAAGGAATGCAGCATGGATCCCTGCCCATCTAGTGTAT	993

[illegible]

```

1864 CAGGCTGTAAACACAGAGCCCTGTCCCCAGGTGGCATGTGGGCTCTTG 1913
      :::::||||| :::::|||||  |||||:::|||||:::|||||
675 pSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValp 692
      |||||:::|||||  |||||:::|||||:::|||||:::|||||
1914 GGGGCTCTCTCAGCTACCTGTGGAGTTGGAATTCAGACCCGAGATGTGT 1963
      ::||| |||  ::||| |||  ::||| |||  ::||| |||
692 heCysSerHisLeuSerArgGluMetAsnGluThrValIleLeuAla 708
      ::||| |||  ::||| |||  ::||| |||  ::||| |||
1964 ACTGCTGCAC.....CCAGGGGAGACCCCTCCCTCCT 1998
      ::|||:::|||||:::|||||:::|||||:::|||||
709 AspGluLeuCysArgGlnProLysProSerThrValGlnAlaCysAsnAr 725
      ::|||:::|||||:::|||||:::|||||:::|||||
1999 GAGGAG...TGCGAGATGAAGAGCCCATGCTTACAGCATGCAATCA 2045
      ::|||:::|||||:::|||||:::|||||:::|||||
725 gPheAsnCysProProAlaTTrpTrpProAlaGlnTrpGlnProCysSera 742
      ::|||:::|||||:::|||||:::|||||:::|||||
2046 GTTTCAGTGCCTCTGCTGGCAGCATTAAGAAATGGCAGCATGTCTCCA 2095
      ::|||:::|||||:::|||||:::|||||:::|||||
742 rgThrCysGlyGlyValGlnLysArgGluValLeuCysLysGlnArg 758
      ::|||:::|||||:::|||||:::|||||:::|||||
2096 GGACTTGTGGGGGGAACCTCAGAACAGAGAGTACCTGTGGCAGCTG 2145
      ::|||:::|||||:::|||||:::|||||:::|||||
759 MetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlase 775
      ::|||:::|||||:::|||||:::|||||:::|||||
2146 CTAACGGATGCGAGCTTTTGAATCTCTCAGATGAATTTGTCAGAGACC 2195
      ::|||:::|||||:::|||||:::|||||:::|||||
775 rLysProAlaCysGlnGlnAlaCysLysLysAspAspCysProSerGluT 792
      ::|||:::|||||:::|||||:::|||||:::|||||
2196 CAAGGCATCGTCTCACAAGTCTGTGCCAGGACAGACGTCTCTCCACATT 2245
      ::|||:::|||||:::|||||:::|||||:::|||||
792 rpLeuLeuSerAspTrpThrGlu 799
      ::|||:::|||||:::|||||:::|||||:::|||||
2246 TAGCTGTGGGAGATGGTGAAG 2268
      ::|||:::|||||:::|||||:::|||||:::|||||

```

seq\_name: gb\_pat:AX235354

seq\_documentation\_block:

LOCUS AX235354 6294 bp DNA linear PAT 11-SEP-2001

DEFINITION Sequence 1 from Patent WO0162928.

ACCESSION AX235354

VERSION AX235354.1 GI:15593887

KEYWORDS

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 6294)

AUTHORS Vernet,C.A., Fernandes,E., Shimkets,R.A., Macdougall,J. and Spaderna,S.K.

TITLE Polypeptides and nucleic acids encoding same

JOURNAL Patent: WO 0162928-A 1 30-AUG-2001;

Curagen Corporation (US)

FEATURES

source

1..6294

/organism="Homo sapiens"

/db\_xref="taxon:9606"

416..4261

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC69811.1"

/db\_xref="GI:1559388"

/translation="MPDYHFPQLPRWHPNPTACSVCGGQIRRSFVCVESMHGEI

LQVBEWCMYAPKPKVMQCNLFCPKWIAMWSQCTVTCGRGLRYRVVLGINRGEH

VGGNPOILKLIKECVITIPCYKPKESPVKEAKLPLKQAELEETRIATPEPTFIP

EPWSACSTTCGPGVQVREVKRVLTTQTTELPEECGPKLPTERPCLLEACDES

PASRELDITPLDESETTYDWEYAGFTPCTATCVGHQBEAIVCLHIQTQQTVPNDLSD

DEKPHALQCNQFCPCPGWHLEWQOCRTCCGGTGNRRVTCRQLITDGSPLNLSDEL

CMCKPASHSKSCARTDPPHILAVGWSKCSVCGVGTQRRKQVQORLAAGKRRILPSE

MMCRDLPLPLRVKSCMPCKSIKSEMTKLGEQGPOLLVSQVYIOTREKRNLTIT

GSRAYLTPNTSVIITKCPVRFQKSLIQWEDKGRQLNSKRLGTFKSGSLKHGLAAPP

IGVYRCIAGSNAQEVTVLKLIGTDNRLIARPALREPMREYPGMDHSEANSLSGVTWHKMR

OMWNKNLDYLDDBDHISNOPELRALLGHCSNSAGSTNSWELKNKQFEAAVKGAYSMD  
TAQDELIRNNSQLMEGEVSDDLASLIYOLVAELAKAQPTHMWRGIOBETPPAAQ  
LRGETGSVSSSHAKNSKULFKPKGPVLMRQSPQSISFNKTIINSRTGNTVYITKRT  
EVINILCDLIITPSATVTWTKDILLQPSVKIILDGTGKIQTNPTRKEOGIYECSSA  
NHLGSDVESSVLVYAEAPVILSVERNITKPEHNLSVVVGGIVAAALGANVTIRCPVK  
GVQPNITWLKRGSLSGNSVLLFNGSILLQNVSLNEGTVVCIATNALKAVATSVL  
HLLERRWPESRIVFLOGHKYILQATNTRTNSNPTGEPPQEFWFEPGNWHSATC  
GHLGARTORPCQVMANGQVSEALCDHLKPLAGFPCNIRDCEPARFTSVNSQCSVS  
CGEGYHSRQVTCRRTKANGTVQVVSPPACAPKPLGRKPCFQHPCYQWEPGNRCPCR  
CMGRAVHMQRHHTACOHNSDNCDDRRKPTLRNCTSGACDVCWHGTGPKPKPTAACG  
RGFOSRKVDCHITRSCKPVAKRHCVKQKKPITSWRHCLGPSRCDRCTDITTHCYMFKHL  
NLCSLDRYKORCCQCOEG\*

BASE COUNT 1789 a 1297 c 1528 g 1680 t  
ORIGIN

alignment\_scores:

Quality: 2597.50 Length: 1435

Ratio: 2.851 Gaps: 31

Percent Similarity: 63.484 Percent Identity: 37.561

alignment\_block:

US-10-044-807-2 x AX235354 ..

Align seg 1/1 to: AX235354 from: 1 to: 6294

```

362 SerAspGlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuPr 378
      |||||:::|||||:::|||||:::|||||:::|||||
395 ACTGATGGATTAAAGAGATAATGCCCTATGACCACCTTCCAACTCTTCC 444
      ::|||:::|||||:::|||||:::|||||:::|||||
378 oArgTrpGluAlaThrProTrpThrAlaCysSerSerCysGlyGlyG 395
      ::|||:::|||||:::|||||:::|||||:::|||||
445 TCGCTGGGAACATAATCTTGGACTGCATGTTCCGTCTCTGTGGAGGAG 494
      ::|||:::|||||:::|||||:::|||||:::|||||
395 lYleGlnSerArgAlaValSerCysValGluGluAspIleGlnHis 411
      ::|||:::|||||:::|||||:::|||||:::|||||
495 GGATTCAGAGACGGAGCTTTGTGTGTAGAGGAATCCATGCATGGAGAG 544
      ::|||:::|||||:::|||||:::|||||:::|||||
412 ValThrSerValGluGluTrpLysCysMetTyrThrProLysMetProil 428
      ::|||:::|||||:::|||||:::|||||:::|||||
545 ATATTGCAGTGGGAAGAATGGAAGTGCATGTACGCCAACCAACCAAGGT 594
      ::|||:::|||||:::|||||:::|||||:::|||||
428 eAlaGlnProCysAsnIlePheAspCysProLysTrpLeuAlaGlnGluT 445
      ::|||:::|||||:::|||||:::|||||:::|||||
595 TATGCAAACTGTAACTGTTGATGCCCAAGTGGATGCCATGGAGT 644
      ::|||:::|||||:::|||||:::|||||:::|||||
445 rpSerProCysThrValThrCysGlyGlnGlyLeuArgTyrArgValVal 461
      ::|||:::|||||:::|||||:::|||||:::|||||
645 GGTCTCAGTCACAGTACTTGTGGCGAGGTTACGGTACCGGGTGTGT 694
      ::|||:::|||||:::|||||:::|||||:::|||||
462 LeuCysIleAspHisArgGlyMetHisThrGlyGlyCysSerProLysTh 478
      ::|||:::|||||:::|||||:::|||||:::|||||
695 CTGTGTATTAAACACCCCGGAGACATGTTGGGGGCTCCAATCCCAACT 744
      ::|||:::|||||:::|||||:::|||||:::|||||
478 rLysProHisIleLysGluGluCysIleValProThrProCysTyrLysp 495
      ::|||:::|||||:::|||||:::|||||:::|||||
745 GAAGTTACATCAACAGAGAATGTGTCATTCCTCCATCCCTGTTATAAAC 794
      ::|||:::|||||:::|||||:::|||||:::|||||
495 rLysGluLysLeuProValGluAlaLysLeuProTrpPheLysGlnAla 511
      ::|||:::|||||:::|||||:::|||||:::|||||
795 CAAAGAAAAAGTCCAGTGGAGCAAAATTCCTTGGCTGGAACAAGCA 844
      ::|||:::|||||:::|||||:::|||||:::|||||
512 GlnGluLeuGluGluGlyAlaAlaValSerGluGluProSerPheIlePr 528
      ::|||:::|||||:::|||||:::|||||:::|||||
845 CAGAAGACTAGAAGACACAGAAATAGCAACAGAAAGAACACGTTCAITCC 894
      ::|||:::|||||:::|||||:::|||||:::|||||
528 oGluAlaTrpSerAlaCysThrValThrCysGlyValGlyThrGlnVala 545
      ::|||:::|||||:::|||||:::|||||:::|||||
895 AGAACCTGTGTAGCTGTGACCTACCACTGGCGCGGTGTGCAGGTCC 944
      ::|||:::|||||:::|||||:::|||||:::|||||
545 rgtIleValArgCysGlnValIleuSerPheSerGlnSerValAlaAsp 561
      ::|||:::|||||:::|||||:::|||||:::|||||
945 GTGAGGTGAAGTCCCGTGTCTCTCACATTACGACACTGAGACTGAGTGAG 994
      ::|||:::|||||:::|||||:::|||||:::|||||

```

```

562 LeuProIleAspGluCysGluGlyProLysProAlaSerGlnArg..... 576
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
995 CTGCCGAGGAAGAGTGTGAAGCCCAAGCTCCACCGAAGCGCCCTG 1044
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 .....AlaCysTyrAlaGlyProCysSerGlyGluIleProGluP 590
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1045 CTTCCTGGAGCATGATGAGAGCCCGGCTCCCGAGAGCTAGACATCC 1094
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 heAsnProAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheAsp 606
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1095 CTCTCCCT.....GAGGACAGTGAG 1114
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
607 GluLeuTyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCy 623
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1115 ACGACTTACCACGCTGGAGTACGCTGGGTTCACCCCTGCACAGCAACATG 1164
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
623 sGlyGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrA 640
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1165 CGTGGGAGGCCATCAGAAGCCATAGCAGTGTGCTTACATATCCAGACCC 1214
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 rgGluProAlaGluGluAsnLeuCysValThrSerArgArgProProGln 656
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1215 AGCAGACAGTCAATGACAGCTTGTGTGATATGTTCCACCGTCCCTCAGCC 1264
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 LeuLeuLysSerCysAsnLeuAspProCysProAlaArgTrpGluIleG1 673
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1265 ATGAGCAGGCGCTGTACACAGACCCCTGTCCCCAGGTGGCATGTGGG 1314
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
673 YLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgA 690
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1315 CTCATTGGGGCCCTGCTCAGTACCTGTGGAGTTGGAATTCAGACCCCGAG 1364
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
690 spValPheCysSerHisLeuLeuSerArgGluMetAsnGluThrValIle 706
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1365 ATGTGTACTGCTGCAC.....CCAGGGGAGACCCCTGCC 1399
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 LeuAlaAspGluLeuCysArgGlnProLysProSerThrValGlnAlaCy 723
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1400 CTCTCTGAGGAG...TGCCGAGATGAAAGCCCAATGCTTTACAGCATG 1446
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
723 sAsnArgPheAsnCysProProAlaTrpTyrProAlaGlnTrpGlnProc 740
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1447 CAATCAGTTTGACTGCTCCCTCTGCTGGTGCACATGAAGAATGGCAGCAGT 1496
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
740 ysSerArgThrCysGlyGlyValGlnLysArgGluValLeuCysLys 756
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1497 GTTCCAGGACTGTGGCGGGGAACCTCAGACAGAGAGTACCTGTGCGG 1546
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 GlnArgMetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSe 773
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1547 CAGCTGCTAACGGATGGCAGCTTTTGAATCTCTCAGATGAATTTGCCA 1596
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
773 rAlaSerLysProAlaCysGlnGlnAlaCysLysLysAspCysProS 790
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1597 AGGACCCAGGCACTGCTCTCAAGTCTGTGCCAGGACAGACTGTCTC 1646
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 erGluTrpLeuLeuSerAspTrpThrGluCysSerThrSerCysGlyGlu 806
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1647 CACATTAGCTGTGGAGACTGCTCGAAGTGTCTGTCASTTGTGTGT 1696
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
807 GlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLe 823
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1697 GGAATCCAGAGAAGAAAGCAGTGTCTCAAGGCTGGCAGCCAAAGCTCG 1746
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
823 uSerThrValValAsnSerThrLeuCysProProLeuProPheSerSerS 840
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1747 GCGCATCCCCCTCAGTGAGATGCTGGCAGGGATCTTACCAGGGCTCCCTC 1796
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
840 erIleArgProCysMetLeuAlaThrCysAlaArg...ProGlyArgPro 855
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1797 TTGTAAGATCTGCCAGATGCTGAGTGCAGTAAATCAATCAGAGATG 1846
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
856 SerThrLys.....HisSerProHisIleAlaAlaArgLysVa 869
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1847 AAGCAAAACTTGTGTCAGAGGGTCCCGAGATCTCAGTGTCCAGAGAT 1896
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
869 lTyrIleGlnThrArgArgGlnArgLysLeuHisPheValValGlyGlyP 886
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1897 CTACATTCAGACAGGAAGAGAGCGTATTAACTGACCATTTGGTAGCA 1946
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
886 heAlaTyrLeuLeuProLysThrAlaValValLeuArgCysProAlaArg 902
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1947 GAGCCTATTGCTGCCCAACACATCCGTGATTAATTAAGTGCCAGTCGGA 1996
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
903 ArgValArgLysProLeuIleThrTrpGluLysAspGlyGlnHisLeuI1 919
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1997 CGAATTCCAGAAATCTGTGATCCAGTGGGAGGATGGCCGTGCCTGCA 2046
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
919 eSerSerThrHisValThrValAlaProPheGlyTyrLeuLysIleHisA 936
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2047 GAATCCNAACGGCTTGGCATCCCAAGTCAGGCTCACTAAATAATCCACG 2096
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
936 rgLeuLysProSerAspAlaGlyValTyrThrCysSerAlaGlyProAla 952
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2097 GTCTTGCTGCCCGCCACATCGCGCTGTACCGGTGCAITTCAGGCTCTGCA 2146
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
953 ArgGluHisPheValIleLeuIleGlyGlyAsnArgLysLeuValAl 969
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2147 CAGAAACAGTTGTGCTCAAGCTCAATGGTACTGACAAACCGGCTCATGCC 2196
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
969 aArgProLeuSerProArgSerGluGluValLeuAlaGlyArgLysG 986
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2197 ACGCCCA.....GCCCTCAGGGAGC 2216
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
986 lyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsnGlyIlePhe 1002
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2217 CTATCAGGGAATATCTCGGATGGACACAGCCCAAGCCCAATAGTTTGGGA 2266
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 SerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnProGlySe 1019
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2267 GTCACATGGCACAAA.....ATGAGGCAATGTGGATAACAACAAATGA 2310
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1019 rArgTyr.....AspAspLeuValSerArg.....L 1028
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2311 CTTTATCTGGATGATGACCATATAGTAACCCAGCCTTTCTTGAGAGCTC 2360
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1028 euLeuGluGlnGlyTrpProGlyGluLeuLeuAlaSerTrpGluAla 1044
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2361 TGTTAGGCCACTGCAGCAATCTTCGAGGAAGCACCAACTCTCTGGGAGTGC 2410
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1045 GlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGlyAlaGl 1061
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2411 AAGAATAAGCAGTTTGAAGCAGCAGTTAAACAA.....GGAGCA... 2449
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1061 uGlnValLeuLeuHisLeuProPheThrMetValThrGluGlnArgArgL 1078
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2450 .....TATAGCTGGATACAGCCAG.....T 2471
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1078 euAspAspIleLeuGlyAsnLeuSerGlnGlnProGlu.....GluLeu 1092
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2472 TTGATCAGCTGATAAGAAACATGAGTCAGCTCATGGAACCCGAGAGGTC 2521
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1093 ArgAspLeuTyrSerLysHisLeuValAlaGlnLeuAlaGlnGluIlePh 1109
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2522 ACGCATGATCTGCGTCCCGACGCTGATATATCAGCTGGTGGCCGAATAGC 2571
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1109 eArgSerHisLeuGluHisGlnAspThrLeuLeuLysProSerGluArgA 1126
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2572 CAAGGCACAGCCCAACACATGACG.....TGGCGGGGTCATCCAGGAAG 2615
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1126 rgThrSerProValThrLeuSerProHisLysHisValSerGlyPheSer 1142
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2616 AGACACCTCT.....GCT 2629
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1143 SerSerLeuArgThrSerSerThrGlyAspAlaGlyGlySerArgAr 1159
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2630 GCTCAGCTCACA.....GGGAAACAGGGAGTGTGCCAAAG 2667
1159 gProHisArgLys.....ProT 1165
2668 CTCGCATGCAAAACACACAGCTGACATTCAGCCGAAAGAGCCTG 2717
1165 hrLeuLeuArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluVal 1181
2718 TTCATCATGAGCAA...AGCCAACTCCCTCAATTTCATTTAATAAACA 2764
1182 Val...ThrHisLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSe 1197
2765 ATAATATCCAGGATGGAAATACAGTATACATTAACAAAGACACAGAGGT 2814
1197 rValLeuLeuHisCysGluAlaIleGlyHisProArgProThrIleSerT 1214
2815 CATCAATATACTGTGTGACCTTATACCCCCAGTGAGCCACATATACAT 2864
1214 rpAlaArgAsnGlyGluGluValGlnPheSerAspArgIleLeuLeuGln 1230
2865 GGACCAAGGATGGAACCTTGTACAGCCCTCAGTAAAAATAATTGGAT 2914
1231 ProAspSerLeuGlnIleLeuAlaProValGluAlaAspValGlyPh 1247
2915 GGAACCTGGGAAGATACAGATACAGAACTCTACAAGAAAGAACAGGCAT 2964
1247 eTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleA 1264
2965 ATATGCAATGTCTGTAGTAAATCATCTGTGTCAGATGTGGAAGTCTT 3014
1264 laValThrLeuAlaGlyLysProLeuValLysThrSerArgMetThrVal 1280
3015 CTGTGCTGTATGCAGAGCACCTGTCTGTTGTTGAAGAAATATC 3064
1281 IleAsnThrGluLysProAlaValThrValAspIleGlySerThrIleLy 1297
3065 ACCAAACACAGCACACCATCTGTCTGTGTGTGGAGCATCGTGA 3114
1297 sThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyValP 1314
3115 GGCAGCCCTTGGACACACAGTCGCAATCCGATGTCCTGTAAAGGTGCC 3164
1314 roGluAlaGluValThrTrpPheArgAsnLysSerLysLeu...GlySer 1329
3165 CTCAGCCTAATAATACTTGTGTGAAGAGAGAGGATCTCTGAGTGGCAAT 3214
1330 ProHisHisLeuHisGluGlySerLeuLeuLeuThrAsnValSerSe 1346
3215 GTTTCCTTGCTTTCAATGGATCCCTGTGTGTCAGAAATGTTCCCTTGA 3264
1346 rAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuT 1363
3265 AAATGAAGGAACCTACGTCGTGATAGCCACCAATGCTTGGAAAGCAG 3314
1363 hrGluSerThrGlnLeuLeuIleLeuAspProProGlnValProThrGln 1379
3315 TGGCAACATCTGTATCCACTTGTGGAAGCAAGATGGCCAGAGAGTAGA 3364
1380 LeuGluAspIleArgAla.....LeuLeuAlaAlaThrGlyPr 1392
3365 ATCGTATTTTCGCAAGGACATAAAAGTACATCTCCAGGCAACCAACAC 3414
1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409
3415 TAGAACCAACACG.....AATG 3431
1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425
3432 ACCCA.....ACAGGAGAACCCCG 3451
1426 ProAsnIleThrTrpPheHisGlyGlnProIleValThrAlaThrGI 1442

```

```

3452 CCT..... 3454
1442 yLeuThrHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459
3454 ..... 3454
1459 euSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAla 1475
3454 ..... 3454
1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpSe 1492
3455 .....CAAGAGCCCTTTTGGCA 3471
1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509
3472 GCCTGGTAACGTGGTCACATTTGTCTGCCACCTGTGTTCATTITGGAGCCC 3521
1509 InGlnProArgLeuArgCysLeuLeu...AsnSerThrGluValAsnPro 1524
3522 GCATTACAGAGCCCCAGTGTGTGATGGCCATGGCAGGAGTGTGTGAG 3571
1525 AlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCysAs 1541
3572 GCCCTGTGTGATCACCTCCAGAAGCCACTGGCTGGTGTGAGCCCTGTAA 3621
1541 nArgArgAspCysProSerArgTrpMetValThrSerTrpSerAlaCysT 1558
3622 CATCGGGAGCTGCCAGCGAGTGTGTTCACAAAGTGTGTGTCACAGTGT 3671
1558 hrArgSerCysGlyGlyGlyValGlnThrArgValThrCysGlnLys 1574
3672 CTGTGCTCTGGGTGAAGATACACACTCGCGAGTGTGACGTGCAAGCGG 3721
1575 LeuLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGI 1591
3722 ACAAAAGCCAAATGGAACCTGTGCAGGTGTGTCTCCAAGAGCATGTGCC 3771
1591 nValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysV 1608
3772 T...AAAGACCGCCTCTGGGAAGAAACCATGTTTGTGTCATCATGTG 3818
1608 alGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCysIleGly 1624
3819 TTCAGTGG.....GAACGAGGAACCGGTGTCTCTGGAGTGTGCATGGC 3862
1625 ProHisLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGI 1641
3863 CGTGCTGTGAGGATGCAGCGCTCACAGCTTGTCAACACACAGCTC 3912
1641 yIleThrLeuProSerGluGlnCysSerAlaLeuProArgProValSerT 1658
3913 T.....GACTCCAACTGTGATGACAGAAAGAGACCCACCTTAA 3950
1658 hrGlnAsnCysTrpSerGluAlaCysSerValHisTrpArgValSerLeu 1674
3951 GAAGGAACATGCACATCAGGCGCTGTGTATGTGTGTGSCACACAGGCC 4000
1675 TrpThrLeuCysThrAlaThrCysGlyAsnTyrGlyPheGlnSerArgAr 1691
4001 TGGAGGCCCTGTACAGCAGCCTGTGGCAGG...GOTTTCACAGTCTCGAA 4047
1691 qValGluCysValHisAlaArgThrAsnLysAlaValProGluHisLeuC 1708
4048 ACTGACTGTATCCACACAGAGAGTGTGCAACCTGTGGCCAAAGAGACACT 4097
1708 ysSerTrpGlyProArgProAlaAsnTrpGlnArgCysAsnIleThrPro 1724
4098 GTGTACAGAAAAAGAAACCAATTTCCTGGCGGCACACTGT...CTTGGGCC 4144
1725 CysGluAsnMetGluCysArgAspThrThrArgTyrCysGluLysVally 1741
4145 TCCTGTGATAGAGCTGCACAGACACACTCACTACTGTATGTGTGTAA 4194

```



```

1606 AGACCAAGGCATCGTCTCACAGTCTCTGCCAGGACACTGTCTCTC 1655
790 e rGluTrpLeuSerAspTrpThrGluCysSerThrSerCysGlyGlu 806
1656 CACATTAGCTGGGAGACATGGTCGAAGTGTCTGTCAGTGTGGTGT 1705
807 GlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLe 823
1706 GGAATCCAGAGAGAAAGCAGGTGTGTCAAAAGGCTGCAGCCAAAGCTCG 1755
823 uSerThrValValAsnSerThrLeuCysProProLeuProPheSerSers 840
1756 GCGCATCCCTCAGTCAGATGATGTCCAGGATCTACCAAGGTTCCCTC 1805
840 e rLeuArgProCysMetLeuAlaThrCysAlaArg...ProGlyArgPro 855
1806 TTGTAAGATCTGCCAGATCCCTGAGTCAGTCAGTAAATCAATCAGAGATG 1855
856 SerThrLys.....HisSerProHisIleAlaAlaAlaArgLysVa 869
1856 AAGACAAAACCTTGGTGAGCAGGGTCCGAGATCCTCAGTGTCCAGAGAT 1905
869 lTyrIleGlnThrArgArgGlnArgLysLeuHisPheValValGlyGlyP 886
1906 CTACATTCAGACAGGAAGAGAGAGCGTATTAACCTACCATGGTGTAGCA 1955
886 heAlaTyrLeuLeuProLysThrAlaValValLeuArgCysProAlaArg 902
1956 GAGCCTATTGCTGCCCAACACATCCGTGATTATTAAGTCCCGCTGCGA 2005
903 ArgValArgLysProLeuIleThrTrpGluLysAspGlyGlnHisLeuIl 919
2006 CGATTCCAGAAATCTCTGATCCAGTGGGAGAGGATGGCGGTTGCCCTGCA 2055
919 eSerSerThrHisValThrValAlaProPheGlyTyrLeuLysIleHisA 936
2056 GAATCCAAACGGCTTGGCATCCCAAGTCAGGCTCACTAAATCCACG 2105
936 rgLeuLysProSerAlaGlyValTyrThrCysSerAlaGlyProAla 952
2106 GTCTGTGCTGCCCGCATCGGGGTGTACCGGTGCATTGCAGGCTTCGCA 2155
953 ArgGluHisPheValIleLysLeuIleGlyClyAsnArgLysLeuValAl 969
2156 CAGGAACAGCTTGTGCTCAAGCTCATTTGGTACTGACACCGGCTCATCGC 2205
969 aArgProLeuSerProArgSerGluGluGluValLeuAlaGlyArgLysG 986
2206 ACCGCCA.....GCCCTCAGGGAGC 2225
986 lyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsnGlyIlePhe 1002
2226 CTATGAGGGAATATCCGGGATGACACACAGCGCAATAGTTTGGGA 2275
1003 SerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnProGlySe 1019
2276 GTCACATGGCACAAA.....ATGAGGCAATGTGGAATAACAAAATGA 2319
1019 aArgTyr.....AspAspLeuValSerArg.....L 1028
2320 CCTTATCTCGATGAGCACCACATAGTACCAGCCCTTCTTGCAGAGCTC 2369
1028 euLeuGluGlnGlyIleTrpProGlyGluLeuAlaSerTrpGluAla 1044
2370 TGTTAGGCCACTGCAGCAATCTGCAGGAGCACCACCACTCCTGGAGTTG 2419
1045 GlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGlyAlaGl 1061
2420 AAGAATAAGCAGTTTGAAGCAGCAGGTTAAACAA.....GGAGCA.. 2458
1061 uGlnValLeuLeuHisLeuProPheThrMetValThrGluGlnArgArgL 1078

```

```

2459 .....TATAGCATGGATACAGCCAG.....T 2480
1078 euAspAspIleLeuGlyAsnLeuSerGlnGlnProGlu.....GluLeu 1092
2481 TTGATGAGCTGATAGAAACATGAGTCAGTCATGGAAACCGGAGAGTCT 2530
1093 ArgAspLeuTyrSerLysHisLeuValAlaGlnLeuAlaGlnGluIlePh 1109
2531 AGCGATGATCTTGGTCCCGCTGATATATCATCTGCTGGCGCGATATAGC 2580
1109 eArgSerHisLeuGluHisGlnAspThrLeuLeuLysProSerGluArgA 1126
2581 CAAGGCACAGCCAAACACATGTCAG.....TGGCGGGCATCCAGGAAG 2624
1126 rgThrSerProValThrLeuSerProHisLysHisValSerGlyPheSer 1142
2625 AGACACCTCCT.....GCT 2638
1143 SerSerLeuArgThrSerSerThrGlyAspAlaGlyGlyGlySerArgAr 1159
2639 GCTCAGCTCAGA.....GGGAAACAGGAGTGTGTCCCAAG 2676
1159 gProHisArgLys.....ProT 1165
2677 CTCGCATGCCAAAACTCAGCAAGCTGCATTCAGCCGAAAGGACCTG 2726
1165 hrIleLeuArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluVal 1181
2727 TTCATCAGGSCAA...AGCCAACCTCCCTCANTTTCATTTAATAAACA 2773
1182 Val...ThrHisLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSe 1197
2774 ATAATTCAGGATTTGGAATACAGTATACATTACAAAAGACAGCAGGT 2823
1197 rValLeuLeuHisCysGluAlaIleGlyHisProArgProThrIleSerT 1214
2824 CATCAATATACCTGTGTGACCTTATTACCCCGAGTGAGGCCACATACAT 2873
1214 rpAlaArgAsnGlyGluGluValGlnPheSerAspArgIleLeuLeuGln 1230
2874 GGACCAAGGATGGAACTTGTTCAGCCCTCAGTAAATAATTTTGGAT 2923
1231 ProAspAspSerLeuGlnIleLeuAlaProValGluAlaAspValGlyPh 1247
2924 GGAACCTGGAAAGATACAGATACAGATTCCTACAGGAAGAAACAGGCAT 2973
1247 eTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleA 1264
2974 ATATGAATGTCTCTAGCTAATCATCTTGTTCAGATGTGAAAAGTTCTT 3023
1264 laValThrLeuAlaGlyLysProLeuValLysThrSerArgMetThrVal 1280
3024 CTGTGCTGTATGAGAGCCACTGTCTCATCTGTCTGTGGAAGAAATATC 3073
1281 IleAsnThrGluLysProAlaValThrValAspIleGlySerThrIleLy 1297
3074 ACCAAACAGAGCACCAACCATCTCTGTGTGTGTGGAGCGCATCGTGA 3123
1297 sThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyValP 1314
3124 GGCAGCCCTTGGAGCAACCGTACAACTCCGATGCTCTGTAAAGAGGTGCC 3173
1314 roGluAlaGluValThrTrpPheArgAsnLysSerLysLeu...GlySer 1329
3174 CTCAGCCTAATAATAACTTGGTTGAGAGAGGAGGATCTCTGAGTGGCAAT 3223
1330 ProHisHisLeuHisGlySerLeuLeuLeuThrAsnValSerSerSe 1346
3224 GTTTCCTTGTCTTTCAATGGATCCCTGTGTGTGCAGAAATGTTTCCCTGA 3273
1346 rAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuT 1363
3274 AAATGAAGGAACCTTACGTCTCTAGTACCAATGCTCTTTGGAAGGACG 3323

```

```

1363 hrGluSerThrGlnLeuLeuLeuLeuAspProProGlnValProThrGln 1379
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3324 TGGCAACATCTGTACTCCACTTGTGTGAGAGAGATGCCAGAGATAGA 3373
1380 LeuGluAspIleArgAla.....LeuLeuAlaAlaThrGlyPr 1392
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3374 ATCGTATTTCTGAAGGACATAAAAGTACATTCCTCCAGCAACCAACAC 3423
1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3424 TAGAACCAACAGC.....AATG 3440
1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425
  ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3441 ACCCA.....ACAGGAGAACCCCG 3460
1426 ProAsnIleThrPheHisGlyGlnProIleValThrAlaThrG1 1442
  ||| CCT..... 3463
1442 yLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459
  3463 ..... 3463
1459 euSerGlyGlySerGlnGlyCluPheSerCysLeuAlaGlnAsnGluAla 1475
  3463 ..... 3463
1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSe 1492
  ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3464 .....CAAGAGCCTTTTGGGA 3480
1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3481 GCCTGTAACGTGTCACATGTTCTGCCACCTGTGGTCATTTGGGAGCCC 3530
1509 lnGlnProArgLeuArgCysLeuLeu...AsnSerThrGluValAsnPro 1524
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3531 GCAFTACAGAACCCAGTGTGTGATGTCCTCAATGGCAGGAAGTGAGTGAG 3580
1525 AlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCysAs 1541
  ||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3581 GCCCTGTGTATCACCTCCAGAGAGCCACTGGCTGGGTTTGAGCCCTGTAA 3630
1541 nArgArgAspCysProSerArgTrpMetValThrSerTrpSerAlaCyst 1558
  ||||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3631 CATCCGGCACTGCCAGCGAGGTGTTTCAACAAGTGTGTGTGCACAGTGCT 3680
1558 hrArgSerCysGlyGlyGlyValGlnThrArgArgValThrCysGlnLys 1574
  :: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3681 CTGTGCTTCCGGTGAAGGATACCAAGTCCAGAGTGGCAGGTGACGTGCAAGCG 3730
1575 LeuLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrG1 1591
  ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3731 ACAAAACCAATGGAACGTGTGACAGTGTGTCTCCAGAGCATGTGCCCC 3780
1591 nValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysV 1608
  ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3781 T...AAAGACCGGCTCTGGGAAGAAAACCACTGTTTGGTCATCCATGTG 3827
1608 alGluTrpAlaPheSerSerTrpClyGlnCysAsnGlyProCysIleGly 1624
  ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3828 TTCAGTGG.....GAACAGGGAACCGGTGCTCTGGACGTGTGATGGG 3871
1625 ProHisLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspG1 1641
  :: ::::: ||| ::::: ||| ::::: ||| ::::: |||
3872 CGTGTGTGAGGATGCAGAGCGTCACACAGCTTGTCAACACACAGCTC 3921
1641 yIleThrLeuProSerGlnCysSerAlaLeuProArgProValSerT 1658
  :: ::::: ||| ::::: ||| ::::: ||| ::::: |||
3922 T.....CACTCCAACTGTGTGATGACAGAAAGAGACCCACTTAA 3959

```

```

1658 hrGlnAsnCysTrpSerGluAlaCysSerValHisTrpArgValSerLeu 1674
  :: ::::: ||| ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: |||
3960 GAGGAACTGCACATCATGGGCGCTGTGTGTGTGTGTGGCACACAGGCGCT 4009
1675 TrpThrLeuCysThrAlaThrCysGlyAsnTyrGlyPheGlnSerArgAr 1691
  ||| ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: |||
4010 TGGAGCGCTGTACAGCAGCTGTGGCAGG...GTTTCCAGTCTCGGAA 4056
1691 gValGluCysValHisAlaArgThrAsnLysAlaValProGluHisLeuC 1708
  :: ::::: ||| ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: |||
4057 AGTCACTGTATCCACACACAGAGAGTTGCAAACTGTGGCCCAAGAGACACT 4106
1708 ysSerTrpGlyProArgProAlaAsnTrpGlnArgCysAsnIleThrPro 1724
  || ::::: ||| ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: |||
4107 GTGTACAGAAAGAAACCAATTCCTGGCGGCACTGT...CTTGGGCCC 4153
1725 CysGluAsnMetGluCysArgAspThrArgTyrCysGluLysValLy 1741
  :: ::::: ||| ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: |||
4154 TCCTGTGATAGAGACTGCACAGACACAACCTCACTACTGTATGTTGTA 4203
1741 sGlnLeuLysLeuCysGlnLeuSerGlnPheLysSerArgCysCysGlyT 1758
  ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: ||| ||||| ::::: |||
4204 ACATCTTAATTGTGTTCTCTAGACCGCTACAAACAAGGTGCTGCCAGT 4253
1758 hrCys 1759
  ::|||
4254 CATGT 4258
seq_name: gb_pr:AF251058
seq_documentation_block:
LOCUS AF251058 1616 bp mRNA linear PRI 15-APR-2001
DEFINITION Homo sapiens clone 2 thrombospondin mRNA, complete cds.
ACCESSION AF251058
VERSION AF251058.1 GI:13625177
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1616)
AUTHORS Mao,Y., Xie,Y., Zhou,Z., Zhao,W., Zhao,S., Wang,W., Huang,Y.,
Wang,S., Tang,R., Chen,X. and Wu,C.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2000) Institute of Genetics, School of Life
Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R.
China
FEATURES
source
1..1616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2"
162..1415
/codon_start=1
/product="thrombospondin"
/protein_id="AAK34948.1"
/db_xref="GI:13625178"
/translation="MAGPLTADFIVKIRNSGADSTVQFIYFQPIIHRWRETDFFPCS
ATCGGYQLTSAECYDLRSNVVQYCHYIPENIKPKLQECNLDPCDPSADGYKQI
MPDYLYHPLRWETATWTATSSCGGIGQSRVSCVEEDIOGHVTSVEBWKCMYTPKM
PIAQCNIIFDCPKWLAQWSPCTVTCGGILRYRVILCIDRHGHTGCSKPKPHIKE
ECIVPTCPYKPKELPVEAKLPWFKAQBLEGAASVESEFPIEAWSACTVTCGVGT
QVRIVRCVLLSFQSVAADLPIDEEGPKPQACYAGPCSGEITPEFNPDETDGLFG
GLQDFEIDYDWEYETGKCSGCGVQEAUVSCLNKQTPRAENLCVTSRRPOLL
KSCNLDPCPARSIDSANACNVLC"
BASE COUNT 386 a 438 c 426 g 366 t
ORIGIN

```

```

alignment_scores:
Quality: 2558.00 Length: 534
Ratio: 5.242 Gaps: 3
Percent Similarity: 91.386 Percent Identity: 87.828

```



## alignment\_block:

US-10-044-807-2 x AF251058 ..

Align seg 1/1 to: AF251058 from: 1 to: 1616

```
214 HisIleArgLeuValLeuLysGlyProAspHisLeuTyrLeuGluThrLy 230
|||||
3 CATATTCGCCCTTGCTTAAAGGTCCTGATCATTATATCTGGAACCAA 52
230 sThrLeuGlnGlyThrLysGlyGluAsnSerLeuSerSerThrGlyThrP 247
|||||
53 AACCTCCAGGGGACTAAGGTGAAACAGTCTCAGCTCCACAGGAAGTT 102
247 heLeuValAspAsnSerSerValAspPheGlnLysPheProAspLysGlu 263
|||||
103 TCCITGTGACAAATTCATGTGGACTTCCAGAAATTTCCAGACAAAGAG 152
264 IleLeuArgMetAlaGlyProLeuThrAlaAspPheIleValLysIleAr 280
|||||
153 ATACTGAGAATGGCTGGACCACTCACAGCAGATTTCATGTCAAGATTGG 202
280 gAsnSerGlySerAlaAspSerThrValGlnPheIlePheTyrGlnProI 297
|||||
203 TAATCGGGCTCCGCTGACATGACAGTCCAGTTCATCTTCATCAACCCA 252
297 leIleHisArgTyrArgGluThrAspPhePheProCysSerAlaThrCys 313
|||||
253 TCATCCACCGATGGAGGAGACGGATTTCCTTCTTGCTCAGCAACCTGT 302
314 GlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAs 330
|||||
303 GGAGGAGGTTATCATCGTGTACATCGCTGAGTGTACGATCTCAGGAGCAA 352
330 nArgValValAlaAspGlnTyrCysHisTyrTyrProGluAsnIleLysP 347
|||||
353 CCGTGTGTTCTGCTGACCAATFACTGTCACTATTACCCAGAGAACATCAA 402
347 roLysProLysLeuGlnGluCysAsnLeuAspProCysProAlaSerAsp 363
|||||
403 CCAAAACCAAGCTTCAGGAGTGCACTTGGATCTCTTGTCAGGCAGTGAC 452
364 GlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTr 380
|||||
453 GGATACAGCAGATCATGCCCTTATGACCTCTACCATCCCTTCCTCGGTG 502
380 pGluAlaThrProTyrThrAlaCysSerSerCysGlyGlyGlyIleG 397
|||||
503 GGAGGCCACCCCATGGACCGGTGCTCTCTCTGCTGTGGGGGGGCATCC 552
397 lnSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
|||||
553 AGAGCCGGGCAGTTTCCTGTGTGTGAGGAGGACATCCAGGGGCGTCACT 602
414 SerValGluGluTyrLysCysMetTyrThrProLysMetProIleAlaGl 430
|||||
603 TCAGTGAAGAGTGGAAATGATCATACCCCTAAGATGACCATCGCGCA 652
430 nProCysAsnIlePheAspCysProLysTyrLeuAlaGlnGluTyrSerP 447
|||||
653 GCCCTGAACAATTTTGACTGCCCTTAATGGCTGGCAGAGGTGGTCTC 702
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
|||||
703 CGTGACAGTGACATGTGGCCAGGCGCTCAGATACCGGTGGTCTCTCTGC 752
464 IleAspHisArgGlyMetHisThrGlyCysSerProLysThrLysPr 480
|||||
753 ATCGACCATCAGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGGCC 802
480 oHisIleLysGluGluCysIleValProThrProCysTyrLysProLysG 497
|||||
803 CCACATAAAGAGGATGCATCGTACCCACTCCCTGCTATTAACCCAAAG 852
```

```
497 luLysLeuProValGluAlaLysLeuProTyrPheLysGlnAlaGlnGlu 513
|||||
853 AGAAACTTCCAGTCAGGCCAAGTTGCCATGGTTCAAAACAAGCTCAAGAG 902
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
|||||
903 CTAGAAAGAGGAGCTGCTGTGTGAGAGGAGCCCTCGTTTCATCCAGAGG 952
530 aTyrSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
|||||
953 CTGGTGGCCCTGCACAGTCACTGTGTGTGGGACCCAGGTGCGAATAG 1002
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563
|||||
1003 TCAGTGGCAGGTGCTCTGCTGCTTCTCTCAGTCCGTGGCTGACCTGCC 1052
564 IleAspGluCysGluGlyProLysProAlaSerGlnArgAlaCysTyrAl 580
|||||
1053 ATTGACGAGTGTGAAGGGCCCAAGCCAGCATCCAGCGTGCCTGTTATGC 1102
580 acLysProCysSerGlyGluIleProGluPheAsnProAspGluThrAspG 597
|||||
1103 AGGCCCATGTCAGCGGGGAAATTCCTGAGTCAACCCAGACGAGCAGATG 1152
597 lyLeuPheGlyGlyLeuGlnAspPheAspGluLeuTyrAspTyrGluTyr 613
|||||
1153 GGCTCTTGGTGGCCCTGCAGGATTTCCAGGAGCTGTATGACTGGAGTAT 1202
614 GluGlyPheThrLysCysSerGluSerCysGlyGlyValGlnGluAl 630
|||||
1203 GAGGGGTTTACCAAGTGTCCGAGTCTGTGGAGGAGGTGTCCAGGAGGC 1252
630 aValValSerCysLeuAsnLysGlnThrArgGluProAlaGluGluAsnL 647
|||||
1253 TGTGTGAGCTGCTTGAACAAACACACTCGGAGGCTGCTGAGGAGAAC 1302
647 euCysValThrSerArgArgProProGlnLeuLeuLysSerCysAsnLeu 663
|||||
1303 TGTGGGTGACCAAGCGCGCGCCACACAGCTCTGAAGTCTGCAATTTG 1352
664 AspProCysProAlaArgTyrGluIle...GlyLysTyrSerProCysSe 679
|||||
1353 GATCCCTGCCAGCAAGAAGAGTAGTACACTCAGCATGGAGCGCTGCAA 1402
679 rLeuThrCysGlyValGlyLeuGlnThrArgAspValPheCysSerHisL 696
|||||
1403 CTTCTTTGTAGCAACCAAGAGCGCTGGCTTCTCATCTCTGCTGTGACC 1452
696 euLeuSer.....Arg 699
|||||
1453 AACTAGCTCTGGGCTAGGGCGAGGTGTCTGCCCTTTATGTTCCACAT 1502
700 GluMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLy 716
|||||
1503 CTGCAAGTGAACCTGGTTGTA..... 1523
716 sProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTyrP 733
|||||
1524 .CCTGATGATCTGAGATCCCATGACTTGTCTCATGTCCTCATGATTCCTT 1572
733 yr 733
1573 AT 1574
```

seq\_name: gb\_pat:AX224815

seq\_documentation\_block:

LOCUS AX224815

DEFINITION Sequence 7 from Patent WO0161011.

ACCESSION AX224815

VERSION AX224815.1 GI:15554912

KEYWORDS

2175 bp

DNA

linear

PAT 10-SEP-2001

```

SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2175)
AUTHORS     Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
            and Sands,A.T.
TITLE       Novel human thrombospondin repeat proteins and polynucleotides
            encoding the same
JOURNAL      Patent: WO 0161011-A 7 23-AUG-2001;
            Lexicon Genetics Incorporated (US)
FEATURES
  source     location/Qualifiers
            I..2175
BASE COUNT   583 a 532 c 575 g 485 t
ORIGIN
alignment_scores:
  Quality: 2425.50      Length: 655
  Ratio: 4.378          Gaps: 2
  Percent Similarity: 84.580 Percent Identity: 63.359
alignment_block:
US-10-044-807-2 x AX224815 ..
Align seg 1/1 to: AX224815 from: 1 to: 2175
22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
190 ACCTCAGAAACACTCGTTCAGATGAGACAAAGATGGCACTGGGATGC 239
38 atrpGlyProtrpSerGluCysSerArgTrpCysGlyGlyGlyAlaSerT 55
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
240 TTGGGGGAGCTGGAGTGAAGTCTCCGGACCTGTGGGGGAGGACATCAT 289
55 yrSerLeuArgArgCysLeuSerSerLysSerCysGluGlyArgAsnIle 71
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
290 ATTCTCTCGGAGATGTTTACTGGAAGGAATGTGAAGGGCAGAACATT 339
72 ArgTyrArgThrCysSerAsnValAspCysProProGluAlaGlyAspPh 88
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
340 CGGTACAGACATGCACATCATGATGCTCCCTCCAGATGAGAAATTT 389
88 eArgAlaGlnClnCysSerAlaHisAsnAspValLysHisGlyGlnP 105
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
390 CAGAGCCAGCAGTGTCTCAGCTACAAATGATGTCCAGTATCAGGGCAT 439
105 heTyrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
440 ACTATGAATGGCTTCCACGATATATATGATCCTGTGCCCCCTGTGCATC 489
122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
490 AAGTGTCTATGCACAGACACAAACCTTGGTGGAGCTGGCCACCTAAGT 539
138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
540 ACTGGATGGAACCTGTTGCAACAGGACTCCTTGGACATGTGTACAGT 589
155 lLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLys 171
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
590 GCATCTGTACGGCAGTGGGTGGCATGCGCAACTGGGAAGCAATGCCAAG 639
172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
640 GAGGACAACTGTGGAGTCTGTGCGGGGATGGCTCCACCTGCAGGCTTGT 689
188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspAspThrV 205
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
690 ACGGGGACAAATCAAAGTCACACGTTTCTCTGAAAAAAGAGAAAAATG 739

```

```

205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
740 TAATTGCTGTTCCCTTTGGGAAGTGAAGTGCAGAAATTACAGTAAGGA 789
222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyCl 238
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
790 CCTGCCCAACCTCTTTATTGAATCAAAACACTTCAAGGAAGCAAGAGGA 839
238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerVala 255
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
840 ACACAGCTTTAACAGCCCCCGGCTTTGTGCTAGAAAACACACAGATGG 889
255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
890 AATTTCAGAGGGCTCCGAGAGGCAAACTTTTAAGATCCAGACCTCTG 939
272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
940 ATGGCTGATTTCATCTTCAAGACCAGGTACACTGCAGCCAAAGACAGCGT 989
288 rValGlnPheIlePheTyrGlnProIleIleHisArgTrpArgGluThrA 305
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
990 GGTTCAGTTCCTTTTACCAGCCCATCAGTCATCAGTGGAGACAACTG 1039
305 spPhePheProCysSerAlaThrCysGlyGlyTyrGlnLeuThrSer 321
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1040 ACTTCTTTCCCTGCACCTGTGAGTGTGAGGAGGTTATCAGCTCAATCT 1089
322 AlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1090 GCTGAATGTGTGATATCCGCTTGAAGAGGGTAGTCTCTGACCATATATG 1139
338 sHisTyrTyrProGluAsnIleLysProLysProLysLeuGlnGlyCysA 355
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1140 TCACCTACTCCCTGAAATGTAAACCAAAACCAAACTGAAGGAATGCA 1189
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1190 GCATGGATCCCTGCCCATCAAGTATGATGATTTAAAGAGATAATGCCCT 1239
372 AspLeuTyrHisProLeuProArgTrpGluAlaThrProTrpThrAlaCy 388
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1240 GACCACTTCCAAACCTTTCCTCGCGGGAACATAATCCTTGACCTGCAIG 1289
388 sSerSerSerCysGlyGlyIleGlnSerArgAlaValSerCysValG 405
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1290 TTCCGTGCTCTGTGAGGAGGATTCAGAGACGGAGCTTGTGTGTGTAG 1339
405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1340 AGGAATCCATGCATGGAGAGATATTGCAGGTGGAAGATGAAGTGCATG 1389
422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1390 TACGCACCCAAACCAAGTTATGCAAACTTGTAAATCTGTTGTATGGCC 1439
438 oLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1440 CAAGTGGATTGCCATGGAGTGTCTCAGTGCACAGTGAAGTGTGGCCGAG 1489
455 lLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1490 GGTACGGTACCGGGTGTCTGTGTATTAACCCACCGCGGAGAGATGTT 1539
472 GlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIleVa 488
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1540 GGGGCTCATTCACAACTGAAGTTACACATCAAGAAAGAAATGTGTCTAT 1589
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
1590 TCCCATCCCGTGTATTAACCAACCAAAAGAAAGTCCAGTGAAGCAAAAT 1639
505 euProTrpPheLysGlnAlaGlnGluLeuGluGlyAlaAlaValSer 521

```



[illegible]

```

1704 TGGCAGCCAAAGGTCGGCGCATCCCCCTCAGTGAGATGATGTGCAGGGAT 1753
      ::      |||      :::::::::::      ::::|||
835  LeuProPheSerSerIleArgProCysMetLeuAlaThrCysAlaAr 851
      |||||      ::::|||      ||::|||      ::::|||
1754 CTACACGGTTCCTCTCTGTGTAGATCTTGCCAGATGCCCTGAGTGCAGTAA 1803
      ::::|||      ::::|||      ::::|||      ::::|||
851 g...ProGlyArgProSerThrLys.....HisSerProHisIleA 864
      ::::|||      ::::|||      ::::|||      ::::|||
1804 AATCAAAATCAGATGAAGACAAAACCTTGGTGAGCAGGGTCGCGACATCC 1853
      ::::|||      ::::|||      ::::|||      ::::|||
864 laAlaAlaArgLysValIleGlnThrArgArgGlnbArgLysLeuHis 880
      ::::|||      ::::|||      ::::|||      ::::|||
1854 TCAGTGCCACGAGACTACATTCAGACAAGGAAGAGAGAGCGTATTAAAC 1903
      ::::|||      ::::|||      ::::|||      ::::|||
881 PheValValGlyGlyPheAlaIleLeuLeuProLysThrAlaValValle 897
      ::::|||      ::::|||      ::::|||      ::::|||
1904 CTGACCATTTGTAGCAGAGCCTATTGTGCGCCCAACACATCCGTGATTAT 1953
      ::::|||      ::::|||      ::::|||      ::::|||
897 uArgCysProAlaArgArgValArgLysProLeuIleThrTrpGluLysA 914
      ::::|||      ::::|||      ::::|||      ::::|||
1954 TAAGTGCCCACTGGCAGATTTCAGAAATCTCTGATCCAGTGGGAGGAGG 2003
      ::::|||      ::::|||      ::::|||      ::::|||
914 spGlyGlnHisLeuIleSerSerThrHisValThrValAlaProPheGly 930
      ::::|||      ::::|||      ::::|||      ::::|||
2004 ATGCCCGTTCCTCGACAACATCCAAAGCGCTTGGCATCAACCAAGTCAGGC 2053
      ::::|||      ::::|||      ::::|||      ::::|||
931 TyrLeuLysIleHisArgLeuLysProSerAspAlaGlyValIleThrCy 947
      ::::|||      ::::|||      ::::|||      ::::|||
2054 TCACATAAAATCCACGGTCTTGTCTGCCCGCGACATCGCGGTGTACCGGTG 2103
      ::::|||      ::::|||      ::::|||      ::::|||
947 sSerAlaGlyProAlaArgGluHisPheValIleLysLeuIleGlyClyA 964
      ::::|||      ::::|||      ::::|||      ::::|||
2104 CATTCAGGCTCTGCACAGGAACAGATTGTGCTCAAGCTCATTTGGTACTG 2153
      ::::|||      ::::|||      ::::|||      ::::|||
964 snArgLysLeuValAlaArgProLeuSerProArgSerGluGluGluVal 980
      ::::|||      ::::|||      ::::|||      ::::|||
2154 ACAACGGCTCATCGCAGCCCA..... 2176
      ::::|||      ::::|||      ::::|||      ::::|||
981 LeuAlaGlyArgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHI 997
      ::::|||      ::::|||      ::::|||      ::::|||
2177 ..GCCCTCAGGGAGCCTATGAGGAATATCTGGGATGGACCACAGCCGA 2223
      ::::|||      ::::|||      ::::|||      ::::|||
997 sGlnAsnGlyIlePheSerAsnGlySerLysAlaGluLysArgGlyLeuA 1014
      ::::|||      ::::|||      ::::|||      ::::|||
2224 AGCCAATAGTTTGGAGTCACATGCGACAAA.....ATGAGCAAAATGT 2267
      ::::|||      ::::|||      ::::|||      ::::|||
1014 laAlaAsnProGlySerArgTyr.....AspAspLeuValSerArg... 1027
      ::::|||      ::::|||      ::::|||      ::::|||
2268 GGAATAACAAAATGACCTTATCTGGATGATGACACACATTAATACCCAG 2317
      ::::|||      ::::|||      ::::|||      ::::|||
1028 .....LeuLeuGluGlnGlyGlyTrpProGlyGluLeuLe 1039
      ::::|||      ::::|||      ::::|||      ::::|||
2318 CCTTTCCTGAGAGCTCTGTAGCCCACTGCACGAATCTGCAGGAAGCAC 2367
      ::::|||      ::::|||      ::::|||      ::::|||
1039 uAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrSerGluG 1056
      ::::|||      ::::|||      ::::|||      ::::|||
2368 CAACTCCTGGGAGTTGAAGATAAAGCAGTTTGAAGCAGCAGTTTAAACAA 2416
      ::::|||      ::::|||      ::::|||      ::::|||
1056 luAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMetVal 1072
      ::::|||      ::::|||      ::::|||      ::::|||
2417 .....GGAGCA.....TATACATGGAT 2434
      ::::|||      ::::|||      ::::|||      ::::|||
1073 ThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGlnGlnPr 1089
      ::::|||      ::::|||      ::::|||      ::::|||
2435 ACAGCCCAAG.....TTTGTAGCTGATAAGAAACATGATGACAGCTCAT 2478
      ::::|||      ::::|||      ::::|||      ::::|||
1089 oGlu.....GluLeuArgAspLeuTyrSerLysHisLeuValAlaGlnI 1104
      ::::|||      ::::|||      ::::|||      ::::|||
2479 GGAAACCGGAGGAGGTCGATGATCTTCGTCGCCAGCTGATATATCAGC 2528
      ::::|||      ::::|||      ::::|||      ::::|||
1104 euAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeu 1120
      ::::|||      ::::|||      ::::|||      ::::|||

```



```

seq_name: gb_pr.AL591423

seq_documentation_block:
LOCUS       AL591423              54193 bp    DNA    linear    PRI 16-NOV-2001
DEFINITION   Human DNA sequence from clone RP11-134P18 on chromosome 9, complete
              sequence.
ACCESSION    AL591423
VERSION      AL591423.6    GI:16973934
KEYWORDS     HTG.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 54193)
DIRECT SUBMISSION
Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 17, 2001 this sequence version replaced gi:16214807.
During sequence assembly data is compared from overlapping clones.
When differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr-9
RP11-134P18 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-134P18 It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-220B22 is at 52194 in this
sequence. The true right end of clone RP11-503K16 is at 2000 in
this sequence.

FEATURES             Location/Qualifiers
     source           1..54193
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-134P18"
                     /clone_lib="RPCI-11.1"
     BASE COUNT      15776 a 11257 c 10727 g 16433 t
     ORIGIN

alignment_scores:
Quality: 2041.00      Length: 723
Ratio: 4.769         Gaps: 2
Percent Similarity: 59.198      Percent Identity: 58.921

alignment_block:
US-10-044-807-2 x AL591423
Align seq 1/1 to: AL591423 from: 1 to: 54193

```

799 GluCysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCy 815  
:::|||||  
24616 CAGTGTTCACAAAGCTGGGGAAGCACCAGACTCGAAGCGCATTTG 24665  
815 sArgLysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuC 832  
|||||  
24666 CCGAAGATGCTGAACACCGGCTCTCAACGGTGTGTCATTTCCACCCCTGT 24715  
832 ysProProLeuProPheSerSerSerIleArgProCysMetLeuAlaThr 848  
|||||  
24716 GCCGGCCCTTCCTTCTCTCCATCAGGCCCTGTATGCTGCAACC 24765  
849 CysAla..... 850  
24766 TGTGCAACTAAGTATGTCAGGGCTCTGGAAATGGGAGATGAACCCACA 24815  
850 ..... 850  
24816 CAGCAGCAGCTATAGCCACCACGCGGTGGCTTCCTTAACCTCAAGACCT 24865  
850 ..... 850  
24866 GTGGGAAATAGTGGACAGTAGAACCCCATGGAGGCTGCAGGAGAGA 24915  
850 ..... 850  
24916 GCTCAGCTGGACAGGAAAGGAATCTTAATTACAGCTGGCCAGGCTAAA 24965  
850 ..... 850  
24966 TTCTCCTCTCCACTGGAACACATTTGCAGGGGGTAGACAAAGTTCA 25015  
850 ..... 850  
25016 CTTGCTGGCATACAGATCTGGAAGCTGGAGGAGTGTATTACTAATGA 25065  
850 ..... 850  
25066 TGTCTAATGTTATCCACTCTGGGGTTTTAAACGACTCTAAACAGATTGT 25115  
850 ..... 850  
25116 TCTGTTAATTAATCCCTAGAGCCGCAACACTTCTAGGCTGTCTTAC 25165  
850 ..... 850  
25166 CATGCTCTTTAGGCTCTGTGCTTGCAGGACCCAGAACCAACATTT 25215  
850 ..... 850  
25216 AGCGGAACACATCTCTCTCTGTCCTCTGCTTTTATGAGGTTGAT 25265  
850 ..... 850  
25266 TTCTGTTTAAAAATATTATTAATGAGAGGAATGCTTCCCGGATC 25315  
850 ..... 850  
25316 CCCTCCATCTAGAACCTGCTATCTCAGTTCTAGGACCAACAGGCCA 25365  
850 ..... 850  
25366 GCAGGGAGACATTTGGGTTGCCACTGCATCTCATTTAGTTCTGATTGGT 25415  
850 ..... 850  
25416 GGCAGCTTGATGGTCCACACGCGGGAATGAGTCTCTCCCGAAACACAG 25465  
850 ..... 850  
25466 GCAGTAAAGGATGCCCGGCCAAAATAAACCCGAGAGATATCTTCTCT 25515  
850 ..... 850

25516 CCGTCTCCGCGACCTTGTCTCTCCCTTCTCTTCTCTCCACTCTGGCTCTT 25565  
850 ..... 850  
25566 TCCTTTGCCCTCTCTCTGGTGCATCTCACTGGGTTTTTCTCTCTCC 25615  
851 ..... Arg...ProGlyA 854  
25616 CCACCTCTTTCTCTCTCGGTTTCTCTCTCTCCAGGGCCCCGGC 25665  
854 rgProSerThrLysHisSerProHisIleAlaAlaAalaArgLysValTyr 870  
|||||  
25666 GGCCATCCACGAAGCACAGCCGCACATCGCGCCGCCAGGAAGGTCTAC 25715  
871 IleGlnThrArgArgGlnArgLysLeuHisPheValValGlyGlyPheAl 887  
|||||  
25716 ATACAGACTCGCAGCAGAGAAAGCTCACTTCGTGGTGGGGGCTTGGC 25765  
887 aTyrLeuLeuProLysThrAlaValValLeuArgCysProAlaArgArgV 904  
|||||  
25766 CTACCTGCTCCCAACAGCGCGGTGGTCTGGCTGCCCGCGCGCAGGG 25815  
904 alArgLysProLeuIleThrTrpGluLysAspGlyGlnHisLeuIleSer 920  
|||||  
25816 TCCGCAAGCCCTCATCCTTGGGAGAAAGGACGGCCAGCACCCTCATCAGC 25865  
921 SerThrHisValThrValAlaProPheGlyTyrLeuLysIleHisArgLe 937  
|||||  
25866 TCGAGCAGCTCACGGTGGCCCTCTCGGCTATCTCAAGATCCACCGCT 25915  
937 uLysProSerAspAlaGlyValTyrThrCysSerAlaGlyProAlaArgG 954  
|||||  
25916 CAAGCCCTCGGATCGAGCGCTTACACCTGCTCAGCGGGCCCGCCGGG 25965  
954 luHisPheValIleLysLeuIleGlyGlyAsnArgLysLeuValAlaArg 970  
|||||  
25966 AGCACTTTGTATTAAAGCTCATCGGAGCAACCGCAAGCTGCTGCGCGG 26015  
971 ProLeuSerProArgSerGluGluValLeuAlaGlyArgLysGlyG 987  
|||||  
26016 CCCGTGAGCCCGAGAAGTGAGAGAGGTGCTTGCGGGGAGGAGCGGG 26065  
987 yProLysGluAlaLeuGlnThrHisLysHisGlnAsnGlyIlePheSerA 1004  
|||||  
26066 CCCGAGGAGGCCCTGCAGACCCACAAACACACAGAGGATCTTCTCCA 26115  
1004 snGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnProGlySerArg 1020  
|||||  
26116 ACGGCAGCAAGCGGAGAAAGCGGCTTGGCGCCCAACCCCGGGAGCGGC 26165  
1021 TyrAspAspLeuValSerArgLeuLeuGlnGlyTyrProGlyG 1037  
|||||  
26166 TACCACGACCTCTGCTCCGCTCTGTGAGAGCGGCGCTGCGCGGAGA 26215  
1037 uLeuLeuAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrs 1054  
|||||  
26216 GCTCTGSCCTCTGTGGAGGGCGCAGGACTCTGCGGAAAGGAAACAGACCT 26265  
1054 erGluGluAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThr 1070  
|||||  
26266 CGGAGGAGGACCCGGTGCAGCAAGTGTCTCTGACCTGCCCTTCACC 26315  
1071 MetValThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGl 1087  
|||||  
26316 ATGTGTACCGAGCAGCGCGCTGGAGACATCTCTGGGAAACCTCTCCA 26365  
1087 ngProGluGluLeuArgAspLeuTyrSerLysHisLeuValAlaGlnL 1104  
|||||  
26366 GCACCCGAGAGCTGCGGACCTCTACAGAACACCTGCTGGCCGAGC 26415  
1104 euAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeu 1120  
|||||

26416 TGGCCAGGAGATCTTCCGAGCCACTGGAGACCAGGACACGCTCCTG 26465  
1121 LysProSerGluArgThrSerProValThrLeuSerProHisIysHi 1137  
|||||  
26466 AAGCCCTGGAGCGCAGGACTTCCCAGTCACTCTCTCGCTCATAAACA 26515  
1137 sValSerGlyPheSerSerSerLeuArgThrSerSerThrGlyAspAlaG 1154  
|||||  
26516 CGTGTCTGGCTTCAGCAGCTCCCTGGGACCTCTCCACCGGGGACGCCG 26565  
1154 lYcGlyGlySerArgArgProHisArgLysProThrIleLeuArgLysIle 1170  
|||||  
26566 GGGGAGGCTCTGAAGGCCACACCGCACCCACCATCTCTGCGCAAGATC 26615  
1171 SerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHisLeuGlyG1 1187  
|||||  
26616 TCAGCGGCCCGCAGCAGCTCTCAGCCTCGGAGGTGTCACCCACCTGGGGCA 26665  
1187 nThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluA 1204  
|||||  
26666 GACGGTGGCCCTGGCCAGCGGACACTGAGTGTCTTCTGCACTGTGAGG 26715  
1204 lalLeGlyHisProArgProThrIleSerTrpAlaArgAsnGlyGluGlu 1220  
|||||  
26716 CCATCGGCCACCCCAAGGCCCTACCATCAGCTGGGCCAGGAATGGAGAAGAA 26765  
1221 ValGlnPheSerAspArg 1226  
|||||  
26766 GTTCAGTTCAGTGACAGG 26783



Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA978791				9466.50	8712.10
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA960468				9464.50	8710.18
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA979226				6176.50	5682.89
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA947482				4797.50	4413.94
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA947458				4775.00	4393.20
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAD14364				3956.50	3632.64
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD17757				3953.50	3626.58
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAD14371				3762.50	3454.17
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD14372				3673.00	3366.41
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA9298232				3612.00	3319.69
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9A06676				3423.00	3149.15
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA1999.DAT:AA9234146				2919.00	2685.19
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA9784532				2919.00	2685.19
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA546142				2919.00	2685.19
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9F63447				2910.00	2677.23
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9F63449				2809.50	2581.41
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD14370				2615.50	2403.38
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD14376				2597.50	2377.58
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD17758				2597.50	2377.57
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA166429				2558.00	2353.67
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD14367				2425.50	2228.76
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA9C75770				2351.00	2164.81
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001B.DAT:AA9F71125				2346.50	2153.23
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD14368				2231.00	2050.87
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9A08462				1930.00	1778.09
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA9C74709				1796.50	1640.51
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD14365				1525.00	1402.90
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA1999.DAT:AA9X7684				1476.00	1360.45
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9F7927				1410.00	1301.03
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA1999.DAT:AA9X7593				1341.00	1238.62
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AAAD14366				1331.00	1225.68
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA936862				1151.50	1043.76
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9A8559				1111.00	1035.40
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2002.DAT:AA9597179				1004.00	909.71
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2002.DAT:AA9597180				999.00	906.71
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9F63449				991.50	897.20
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001A.DAT:AA9F63447				975.00	882.28
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001B.DAT:AA9592052				947.00	855.41
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA2000.DAT:AA9C75598				947.00	855.41
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001B.DAT:AA9C75598				941.00	846.60
/SIDSL/gcgdata/hold-geneseq-geneseq-emb1/NA20001B.DAT:AA9C75598				941.00	846.60

alignment\_block:

US-10-044-807-2 x AAF97891 ..

Align seg 1/1 to: AAF97891 from: 1 to: 5720

```
1 MetGluCysCysArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17
|||||
67 ATGGAATGCTGCCGTCGGCAACTCCCTGGCACACTGCTCCTCTTCTGGC 116
|||||
17 aPheLeuLeuSerSerArgThrAlaAtrSerGluGluAspArgAspG 34
|||||
117 TTTCTCTGCTCCTGAGTTCCAGAGCCGACGCTCCGAGGAGAGCCGGAGC 166
|||||
34 lyLeuTrpAspAlaTrpGlyProThrSerGluCysSerArgThrCysGly 50
|||||
167 GCGTATGGAGTGGTGGGCCCATGGAGTGAATGCTCACGCACTGGCGG 216
|||||
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysGl 67
|||||
217 GGTGGGCGCTCCTACTCTGAGCGCTGCTGAGCAGCAAGAGCTGTGA 266
|||||
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
267 AGGAAGAAATATCCGATACAGAACATCGATATGTGACTGCCACAG 316
|||||
84 luAlaGlyAspPheArgAlaGlnCysSerAlaHisAsnAspValLys 100
|||||
317 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 366
|||||
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
|||||
367 CACCATGGCCAGTTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 416
|||||
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
417 CCGATGTTCACTCAAGTCCCAAGCCAAAGCAACACCTGGTTGTTGAC 466
|||||
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
467 TAGCACCTAAGTCTTAGATGGTACGGCTTGTATACAGATCTTTGGAT 516
|||||
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuGl 167
|||||
517 ATGTGCATCAGTGGTTATGCAAAATGTTGGCTGCGATCCACAGCTGGG 566
|||||
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
|||||
567 AAGCACCGTCAAGAGATACCTGTTGGGCTGCAACGGAGATGGGTCCA 616
|||||
184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
|||||
617 CCTGCCGCTGGTCCGAGGCGAGTATAATCCCAAGCTCTCGGCAACCAA 666
|||||
201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
|||||
667 TCGGATGATACCTGGTTGCAATTCCTCATGGAAGTAGACATATTCGGCT 716
|||||
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
|||||
717 TGTCTTAAAGGTCTCTGATCATTATATCTGGAACCAACCAACCTCCAGG 766
|||||
234 lyThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250
|||||
767 GGACTAAAGGTGAACACAGTCTCAGCTCCACAGGAACCTTCTCTTGTGGAC 816
|||||
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
|||||
817 AATTCTAGTGGACTTCCAGAAATTTCCACAGCAAAAGAGATACAGAAAT 866
|||||
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
867 GGCTGGACCACTCACAGCAGATTTTCATTTGTCAGATTCGTAACTCGGCT 916
|||||
```

```
284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
|||||
917 CCGCTGCAGTACAGTCCAGTTCATCTTCTATCAACCCATCATCCACCGA 966
|||||
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyTy 317
|||||
967 TGGAGGGAGACGGATTTCTTCTTGTCTCAGCACTGTGGAGGAGTTA 1016
|||||
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
|||||
1017 TCAGCTGACATCGGCTGAGTGTCTAGTCTGAGGAGCAACCGTGGTTG 1066
|||||
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1067 CTGACCAATACTGTCTCATTATACCCAGAGAACATCAACCCCAACCCAA 1116
|||||
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysGl 367
|||||
1117 CTTCAGGAGTGCAGCTTGGATCTTGTCCAGCC..... 1149
|||||
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTyrGluAlaThrP 384
|||||
1150 .....AGTGGGAGGCCACCC 1165
|||||
384 roTrpThrAlaCysSerSerCysGlyGlyIleGlnSerArgAla 400
|||||
1166 CATGCACCGCTGCTCTCTCTGTTGGGGGGCATCCAGAGCCGGCA 1215
|||||
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluGl 417
|||||
1216 GTTTCCTGTGGAGGAGACATCCAGGGGCATGTCATTCACTGAGTGAAGA 1265
|||||
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1266 GTGGAAATGCATGTACACCCCTAAGATGCCATCGCGCAGCCCTGCAACA 1315
|||||
434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
|||||
1316 TTTTGTGACTGCCCTAAATGGCTGGCACAGGAGTGGTCTCCGTGCACAGTG 1365
|||||
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1366 ACATGTGGCCAGGGCCTCAGATACCGTGTGGTCTCTGATCGACCATCG 1415
|||||
467 gGlyMethIsthrGlyGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1416 AGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGCCCAACATAAAG 1465
|||||
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1466 AGGAATGCATCGTACCCACTCCTCTATATAACCCCAAGAGAACTTCCA 1515
|||||
501 ValGluAlaLysLeuProTrpPheLysGlnAlaGlnGluLeuGluGluGl 517
|||||
1516 GTCGAGGCCAAGTTCCTCATGGTTCAACAAAGCTCAAGAGCTAGAAGAAG 1565
|||||
517 yAlaAlaValSerGluGluProSerPheIleProGluAlaTrpSerAlaC 534
|||||
1566 AGCTGCTGTGCAGAGGAGCCCTGTTTCATCCCAAGGCCCTGGTCGGCT 1615
|||||
534 ysThrValThrCysGlyValGlyThrGlnValArgIleValArgCysGln 550
|||||
1616 GCACAGTCACTGTGCTGTGGGGACCCAGGTGCGAATAGTCAGGTGCCAG 1665
|||||
551 ValLeuLeuSerPheSerGlnSerValAlaAspLeuProIleAspGluCy 567
|||||
1666 GTGCTCTCTCTCTCTCAGTCCGCTGCTGACCTGCCCTATTGACGAGTG 1715
|||||
567 sGluGlyProLysProAlaSerGlnArgAlaCysTyrAlaGlyProCysS 584
|||||
1716 TGAAGGGCCCAAGCCAGCATCCAGCGTCCCTGCTTATGTCAGGCCCATGCA 1765
|||||
```

2666	GGGGCTTCGCTTACCTGCTCCCAAGACGGGGTGGTGGCTCGCGTGCCCG	2715
901	AlaArgArgValArgLysProLeuIleThrTrpGluLysAspGlyGlnHis	917
2716	GGCGCAGGGTCCGACAGCCCTCATCACTGGGAGAAGACGCGCCACGA	2765
917	stLeuIleSerSerThrHisValThrValAlaProPheGlyTyrLeuLysI	934
2766	CCTCATCAGCTCGACGCAGCTCACGGTGGCCCTTCGGGTATCTCAAGA	2815
934	leHisArgLeuLysProSerAspAlaGlyValTyrThrCysSerAlaGly	950
2816	TCACCGCCTCAAGCCCTCGATGCGAGCGCTACACTGCTTCAGCGGGC	2865
951	ProAlaArgGluHisPheValIleLysLeuIleGlyGlyAsnArgLysLe	967
2866	CGGGCCGGGAGACATTGTGATTAAGCTATCGAGGCAACCGCAAGCT	2915
967	vAlaAlaArgProLeuSerProArgSerGluGluGluValLeuAlaGlyA	984
2916	CGTGGCCGGCCCTTGAGCCCGAGAAGTGAGAAAGGTGCTTCGGGGGA	2965
984	rgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsnGly	1000
2966	GGAAGGGCGCCCGAAGGAGGCCCTGCAGACCCACAACACACAGAACGG	3015
1001	IlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnPr	1017
3016	ATCTTCTCCACGGCAGCAAGCGGAGAGCGGGCTTGGCCGCCAACCC	3065
1017	oGlySerArgTyrAspAspLeuValSerArgLeuLeuGluGlnGlyGlyT	1034
3066	GGGASCCGCTACGAGACCTGCTCTCCGGCTGCTGGAGCAGGGCGGCT	3115
1034	rpProGlyGluLeuLeuAlaSerTrpGluAlaGlnAspSerAlaGluArg	1050
3116	GGCCCGAGAGCTGCTGGCGCTCGTGGGAGGCGAGACTCTGCGGAAGG	3165
1051	AsnThrThrSerGluGluAspProGlyAlaGluGlnValLeuLeuHisLe	1067
3166	AACACAGCTTCGGAGGAGACCCGGGTGCAGAGCAAGTCTCTCGACCT	3215
1067	uProPheThrMetValThrGluGlnArgArgLeuAspAspIleLeuGlyA	1084
3216	GCCCTTACCATGGTGACGACGCGCGGCTTGGACGACATCTCTGGGA	3265
1084	snLeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeu	1100
3266	ACCTCTCCACGACGCCGAGAGACTGCGCGACTCTACAGCAAGCAGCTG	3315
1101	ValAlaGlnLeuAlaGlnIlePheArgSerHisLeuGluHisGlnAs	1117
3316	GTGGCCAGCTGCCCGAGAGATCTCCGACGCCACCTGGAGCACCAGGA	3365
1117	pThrLeuLysProSerGluArgThrSerProValThrLeuSerP	1134
3366	CAGCTCTTGAAGCCCTCGGAGCGCAGACATCCCCAGTGACTCTCTCGC	3415
1134	roHisLysHisValSerGlyPheSerSerSerLeuArgThrSerSerThr	1150
3416	CTCATAAACAGCTGTCTGGCTTTCAGCAGCTCCCTGCGGACCTCTCCACC	3465
1151	GlyAspAlaGlyGlyCysSerArgArgProHisArgLysProThrIleLe	1167
3466	GGGACGCCCGGGGAGGCTCTCGAAGGGCCACCGCAAGCCGACCATCCT	3515
1167	uArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrH	1184
3516	GCGCAAGATCTCAGCGGCCCGACGCTCTACGCTCTGGAGGTGGTCAACC	3565
1184	isLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeu	1200

3566 ACCTGGGCGACAGCGTGGCCCTGGCCAGCGGAGACATGAGTGTCTTCTG 3615  
1201 HisCysGluAlaIleGlyHisProArgProThrIleSerTrpAlaArgAs 1217  
3616 CACTGTGAGGCCATCGGCCACCCCAAGSCCTACCATCAGCTGGGCCAGGAA 3665  
1217 nGlyGluGluValGlnPheSerAspArgIleLeuLeuGlnProAspAsps 1234  
3666 TGGAGAAAGAGTTTCAGTTCAGTACAGGATTCCTACAGCCAGATGATT 3715  
1234 erLeuGlnIleLeuAlaProValGluAlaAspValGlyPheThrCys 1250  
3716 CTTACAGATCTTGGCCACAGTGGAGCAGATGTGGTTTCTACACTTGC 3765  
1251 AsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValThrLe 1267  
3766 AATGCCACCAATGCCITGGGATACGACTCTGTCTCCATTGCCGTACATT 3815  
1267 uAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThrG 1284  
3816 AGCAGGAAAGCCCACTAGTGAACACGTCACGAATGACAGTGATCAACACGG 3865  
1284 luyProAlaValThrValAspIleGlySerThrIleLysThrValGln 1300  
3866 AGAAGCCTGCAGTCACAGTCGATATAGGAGCACCACATCAAAACAGTGCAG 3915  
1301 GlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317  
3916 GGAGTGAATGTGACAATCAACTGCCAGGTGACAGAGTGCTGAAGGTGA 3965  
1317 uValThrTrpPheArgAsnLysSerLysLeuGlySerProHisHisLeuH 1334  
3966 AGTCACCTTGGTTCAGGAATAAAGCAAACTGGGCTCCCGCACCACTGTC 4015  
1334 isGluGlySerLeuLeuLeuThrAsnValSerSerAspGlnGlyLeu 1350  
4016 ACGAAGCTCCTTGTGCTCACAAAGTGTCTCTCGGATCAGGSCCTG 4065  
1351 TyrSerCysArgAlaAlaAsnLeuHisGlyLeuThrGluSerThrG 1367  
4066 TACTCTGGAGGGCGGCCAATCTCAATGAGAGAGCTGACTGAGAGCACCCA 4115  
1367 nLeuLeuIleLeuAspProGlnValProThrGlnLeuGluAspIleA 1384  
4116 GCTGTGATTCCTAGATCCCCCAAGTCCCCACACAGATTGGAAAGACATCA 4165  
1384 rGAlaLeuLeuAlaIaThrGlyProAsnLeuProSerValLeuThrSer 1400  
4166 GGGCCTTGTCTGCTGCCACTGGACCGAACCTTCTTCACTGCTGACGTCT 4215  
1401 ProLeuGlyThrGlnLeuValLeuAspProGlyAsnSerAlaLeuLeuG 1417  
4216 CCTCTGGGAACACAGCTGTGCTCTGGATCTGGGAATTCCTCTCTCTGG 4265  
1417 yCysProIleLysGlyHisProValProAsnIleThrTrpPheHisGlyG 1434  
4266 CTGCCCATCAAGGTCACCTGTCCCTAAATACCTGCTGTTTCATGGTG 4315  
1434 lGlnProIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaIa 1450  
4316 GTCAGCAATTTGCTATGTCACAGGACTGACGCATCACAATCTTGGCAGCT 4365  
1451 GlyGlnIleLeuGlnValAlaAsnLeuSerGlySerGlnGlyGluPh 1467  
4366 GGACAGATCCTTCAAGTTGCAACACCTTACGGTGGGTCTCAAGGGGAATT 4415  
1467 eSerCysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerL 1484  
4416 CAGCTGCCTTGTCTCAGAAATGAGCGAGGGGTGCTCATGACAGAGGCATCTT 4465  
1484 euValIleGlnAspTyrTrpTrpSerValAspArgLeuAlaThrCysSer 1500  
4466 TAGTGATCCAAAGATTACTGGTGGTGTGTGGACAGACTGGCAACCTGCTCA 4515

1501 AlaSerCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLe 1517  
4516 GCCTCCTGTGTGTAAACGGGGGTTTCAGACGCCCGCTTGAGTGCCTGCT 4565  
1517 uAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProA 1534  
4566 GACAGCACGAGGTCAACCTGCCACTGCCAGAGGAAGTTCCGCCCTG 4615  
1534 laValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMet 1550  
4616 CGGTGACAGCCCATCGCTGCAACCCGGAGAGACTGCCCTTCTCGTGGATG 4665  
1551 ValThrSerTrpSerAlaCysThrArgSerCysGlyGlyGlyValGlnTh 1567  
4666 GTGACCTCCTGGTCTGCTGTATCCCGGAGCTGTGGGGAGGTGTCCAGAC 4715  
1567 rArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProv 1584  
4716 CCGCAGGTGACCTGTCAAAAGCTGAAAGGCTCTGGGATCTCCACCCCTG 4765  
1584 alSerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGln 1600  
4766 TGTCCAATGACATGTGACCCAGGTGCGCAAGCGGCTGTGGACACCCAG 4815  
1601 AlaCysAsnGlnGlnLeuCysValGluTrpAlaPheSerSerTrpGlyG 1617  
4816 GCCTGTAAACACGACGCTGTGTGGAGTGGGCTTCTCCAGCTGGGGCCA 4865  
1617 nCysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgLny 1634  
4866 GTGCATGGGCTTGCATCGGCTCACCTAGCTGTGCAACACACAGACAAG 4915  
1634 alPheCysGlnThrArgAspGlyIleThrLeuProSerGluGlnCysSer 1650  
4916 TCTTCTGCCAGACACGGGATGGCATCACCTTACCATCAGACGAGTCAGT 4965  
1651 AlaLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysse 1667  
4966 GCTCTTCCGAGGCTGTGAGCACCAGAACCTGCTGGTTCAGAGCCTCCAG 5015  
1667 rValHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyA 1684  
5016 TGTACACTGGAGAGTCACTGCTGGACCCCTGTGCACAGCTACCTGTGGCA 5065  
1684 snTyrGlyPheGlnSerArgArgValGluCysValHisAlaArgThrAsn 1700  
5066 ACTACGGCTTCCAGTCCCGCGGTGTGGAGTGTGTCATGCCCGCACCAAC 5115  
1701 LysAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTr 1717  
5116 AAGGCAGTGTGCTGAGCACCTGTCTCTGGGGCCCGGCTGCCAACTG 5165  
1717 pGlnArgCysAsnIleThrProCysGluAsnMetGluCysArgAspThrT 1734  
5166 GCAGCGCTGCACATCACCCCATGTGAAACATGGAGTGCAGAGACACCA 5215  
1734 hrArgTyrCysGluLysValLysGlnLeuLysLeuCysGlnLeuSerGln 1750  
5216 CCAGGTACTGCGAGAGGTGAACAGCTGAACCTCTGCCAACTCAGCCAG 5265  
1751 PheLysSerArgCysCysGlyThrCysGlyLysAla 1762  
5266 TTTTAAATCTCGTCTGTGGAACTGTGGCAAGCG 5301

seq\_name: /SIDS1/gcgdata/hold- geneseq/geneseq-emb1/NA2001A.DAT:ABA06468

seq\_documentation\_block:

ID ABA06468 standard; cDNA; 5769 BP.

XX ABA06468;

DT 10-JAN-2002 (first entry)

XX Human cDNA SEQ ID NO: 134.  
XX Human; gene therapy; neural disorder; immune system disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
XX Proliferative disorder; inflammation; ss.  
OS Homo sapiens.  
XX WO200154474-A2. *Prubing*  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01349.  
XX 31-JAN-2000; 2000US-179065P.  
XX 04-FEB-2000; 2000US-180828P.  
XX 24-FEB-2000; 2000US-184664P.  
XX 02-MAR-2000; 2000US-186350P.  
XX 16-MAR-2000; 2000US-189874P.  
XX 17-MAR-2000; 2000US-190076P.  
XX 18-APR-2000; 2000US-198123P.  
XX 19-MAY-2000; 2000US-205515P.  
XX 07-JUN-2000; 2000US-209467P.  
XX 28-JUN-2000; 2000US-214886P.  
XX 30-JUN-2000; 2000US-215135P.  
XX 07-JUL-2000; 2000US-216647P.  
XX 07-JUL-2000; 2000US-216880P.  
XX 11-JUL-2000; 2000US-217487P.  
XX 11-JUL-2000; 2000US-217496P.  
XX 14-JUL-2000; 2000US-218290P.  
XX 26-JUL-2000; 2000US-220963P.  
XX 26-JUL-2000; 2000US-220964P.  
XX 14-AUG-2000; 2000US-224518P.  
XX 14-AUG-2000; 2000US-224519P.  
XX 14-AUG-2000; 2000US-225113P.  
XX 14-AUG-2000; 2000US-225214P.  
XX 14-AUG-2000; 2000US-225266P.  
XX 14-AUG-2000; 2000US-225267P.  
XX 14-AUG-2000; 2000US-225268P.  
XX 14-AUG-2000; 2000US-225270P.  
XX 14-AUG-2000; 2000US-225447P.  
XX 14-AUG-2000; 2000US-225757P.  
XX 14-AUG-2000; 2000US-225758P.  
XX 14-AUG-2000; 2000US-225759P.  
XX 18-AUG-2000; 2000US-226279P.  
XX 22-AUG-2000; 2000US-226681P.  
XX 22-AUG-2000; 2000US-226868P.  
XX 23-AUG-2000; 2000US-227182P.  
XX 30-AUG-2000; 2000US-227009P.  
XX 01-SEP-2000; 2000US-228924P.  
XX 01-SEP-2000; 2000US-229343P.  
XX 01-SEP-2000; 2000US-229344P.  
XX 01-SEP-2000; 2000US-229345P.  
XX 05-SEP-2000; 2000US-229509P.  
XX 05-SEP-2000; 2000US-229513P.  
XX 06-SEP-2000; 2000US-230437P.  
XX 08-SEP-2000; 2000US-231242P.  
XX 08-SEP-2000; 2000US-231243P.  
XX 08-SEP-2000; 2000US-231244P.  
XX 08-SEP-2000; 2000US-231413P.  
XX 08-SEP-2000; 2000US-231414P.  
XX 08-SEP-2000; 2000US-232080P.  
XX 12-SEP-2000; 2000US-232081P.  
XX 14-SEP-2000; 2000US-232397P.  
XX 14-SEP-2000; 2000US-232398P.  
XX 14-SEP-2000; 2000US-232399P.  
XX 14-SEP-2000; 2000US-232400P.  
XX 14-SEP-2000; 2000US-232401P.  
PR 14-SEP-2000; 2000US-233063P.  
PR 14-SEP-2000; 2000US-233064P.  
PR 14-SEP-2000; 2000US-233065P.  
PR 21-SEP-2000; 2000US-234223P.  
PR 25-SEP-2000; 2000US-234274P.  
PR 25-SEP-2000; 2000US-234997P.  
PR 26-SEP-2000; 2000US-234998P.  
PR 27-SEP-2000; 2000US-235484P.  
PR 27-SEP-2000; 2000US-235834P.  
PR 27-SEP-2000; 2000US-235836P.  
PR 29-SEP-2000; 2000US-236327P.  
PR 29-SEP-2000; 2000US-236367P.  
PR 29-SEP-2000; 2000US-236368P.  
PR 29-SEP-2000; 2000US-236369P.  
PR 29-SEP-2000; 2000US-236370P.  
PR 02-OCT-2000; 2000US-236802P.  
PR 02-OCT-2000; 2000US-237037P.  
PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 02-OCT-2000; 2000US-237040P.  
PR 13-OCT-2000; 2000US-239935P.  
PR 13-OCT-2000; 2000US-239937P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241221P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241786P.  
PR 20-OCT-2000; 2000US-241787P.  
PR 20-OCT-2000; 2000US-241808P.  
PR 20-OCT-2000; 2000US-241809P.  
PR 20-OCT-2000; 2000US-241826P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 08-NOV-2000; 2000US-246474P.  
PR 08-NOV-2000; 2000US-246475P.  
PR 08-NOV-2000; 2000US-246476P.  
PR 08-NOV-2000; 2000US-246477P.  
PR 08-NOV-2000; 2000US-246478P.  
PR 08-NOV-2000; 2000US-246523P.  
PR 08-NOV-2000; 2000US-246524P.  
PR 08-NOV-2000; 2000US-246525P.  
PR 08-NOV-2000; 2000US-246526P.  
PR 08-NOV-2000; 2000US-246527P.  
PR 08-NOV-2000; 2000US-246528P.  
PR 08-NOV-2000; 2000US-246532P.  
PR 08-NOV-2000; 2000US-246609P.  
PR 08-NOV-2000; 2000US-246610P.  
PR 08-NOV-2000; 2000US-246611P.  
PR 08-NOV-2000; 2000US-246613P.  
PR 17-NOV-2000; 2000US-249207P.  
PR 17-NOV-2000; 2000US-249208P.  
PR 17-NOV-2000; 2000US-249209P.  
PR 17-NOV-2000; 2000US-249210P.  
PR 17-NOV-2000; 2000US-249211P.  
PR 17-NOV-2000; 2000US-249212P.  
PR 17-NOV-2000; 2000US-249213P.  
PR 17-NOV-2000; 2000US-249214P.  
PR 17-NOV-2000; 2000US-249215P.  
PR 17-NOV-2000; 2000US-249216P.  
PR 17-NOV-2000; 2000US-249217P.  
PR 17-NOV-2000; 2000US-249218P.  
PR 17-NOV-2000; 2000US-249244P.  
PR 17-NOV-2000; 2000US-249245P.  
PR 17-NOV-2000; 2000US-249264P.  
PR 17-NOV-2000; 2000US-249265P.  
PR 17-NOV-2000; 2000US-249297P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 17-NOV-2000; 2000US-249300P.  
PR 01-DEC-2000; 2000US-250160P.  
PR 01-DEC-2000; 2000US-250391P.  
PR 05-DEC-2000; 2000US-251030P.  
PR 05-DEC-2000; 2000US-251988P.  
PR 05-DEC-2000; 2000US-256719P.  
PR 06-DEC-2000; 2000US-251479P.  
PR 08-DEC-2000; 2000US-251856P.

```

PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
DR P-PSDB; ABB10246.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX Claim 1; SEQ ID NO: 134; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
XX Sequence 5769 BP; 1299 A; 1695 C; 1628 G; 1146 T; 1 other;
SO

```

alignment\_scores:                      Quality: 9464.50                      Length: 1762  
   Ratio: 5.427                                      Gaps: 1  
Percent Similarity: 98.978              Percent Identity: 98.978

alignment\_block:  
US-10-044-807-2 x ABA06468    ..

Align seq 1/1 to: ABA06468    from: 1 to: 5769

```

1 MetGluCysCysArgAlaThrProGlyThrLeuLeuLeuAl 17
65 ATGGAATGCTCGCGTGGGCAACTCCTGGCACACTGCTCTTCIGGC 114
17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgasp 34
115 ITTCCTGCTCCTGAGTTCACAGGACCGCACCGCTCCGAGGAGCGCGGACG 164
34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysgly 50
165 GCCTATGGGATGCTGGGGCCCATGGAGTGAATGCTCAGCAGCTGGCGG 214
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG 67
215 GGTGGGCGCTCCTACTCTCTGAGCGCTGCTGAGCAGCAAGAGCTGTGA 264
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
265 AGAAGAAATATCCGATACAGACATGCAGTAATCTGGACTGCCACACG 314
84 luAlaGlyAspPheArgAlaGlnCysSerAlaHisAsnAspVallys 100
315 AAGCAGTGATATCCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 364
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
365 CACCATGGCCAGTTTATGAATGGCTTCCTGTGTCTAATGACCCGTGACA 414
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValGluL 134
415 CCCATGTTCACTCAAGTGCCCAAGCAACCAACCCCTGGTGTGTAAC 464
134 euAlaProLysValLeuAspGlyThrArgCystTyrThrGluSerLeuAsp 150

```

```

465 TAGCACCTAAGGTCTTAGATGGTACGGCTTCTATACAGAAATCTTTGGAT 514
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuGl 167
515 ATGTGCATCAGTGTGTTATGCCAAATGTTGGCTCGGATCACCAGCTGG 564
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlyserT 184
565 AAGCACCGTCAAGGAAGATAACTGTGGGGTCTGCAACGGAGATGGTCCA 614
184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnIleSerAlaThrLys 200
615 CCTGCCGCTGTGTCAGGGCAGTATAATCCAGCTCTCCGCAACCAAAA 664
201 SerAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
665 TCGCATCATCTGTGGTTGCCAATTCCTATGGAAGTAGACATATTCGCT 714
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
715 TGTCTTAAAGGCTCTGATCATCTTATATCTGAAACCAAAACCCCTCCAG 764
234 lyThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250
765 GGACTAAGGTGAACAGCTCTCAGCTCCACAGGAACCTTCCTTGTGGAC 814
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
815 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATACAGAAT 864
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlys 284
865 GGCTGGACCACTCACAGCAGATTTCAATGTCAAGATTGTAACCTCGGCT 914
284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
915 CCGCTGACAGTACAGTCCAGTTCATCTCTATCAACCATCATCCACGA 964
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTY 317
965 TGGAGGAGACGGATTTCTTCTTGTCTCAGCAACCTGTGGAGGAGGTTA 1014
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
1015 TCAGCTGACATCGGCTGAGTCTACGATCTCAGGAGCAACCGTGTGGFTG 1064
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
1065 CTGACCAATAGTCTACTATTACCCAGAGAATCAAAACCAACCCCAAG 1114
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysGl 367
1115 CTTCAGGAGTGCAACTTGGATCTTGTCCAGCC..... 1147
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTrpGluAlaThrP 384
1148 .....AGGTGGAGGCCACCC 1163
384 roTrpThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
1164 CATGACCGGTGCTCCTCTCTGTGGGGGGGCATCCAGAGCGGGCA 1213
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluGl 417
1214 GTTTCCTGTGTGGAGGAGGACATCCAGGGCATGTCACTTCAGTGAAGA 1263
417 uTriPlysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
1264 GTGAAATGCAATGATACACCCCTAAGATGCCATCCGAGCCCTGCAACA 1313
434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
1314 TTTTGTACTGCCCTAAATGGCTGGCACAGGAGTGTCTCCGTGCACAGTG 1363

```

451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467  
|||||  
1364 ACATGTGGCCAGGCGCTCAGATACCGTGTGGTCTCTGCATCGACCATCG 1413  
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484  
|||||  
1414 AGGAATGCACACAGGAGGCTGTAGCCCAAAACAAAGCCACATAAAG 1463  
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500  
|||||  
1464 AGNAATGCATCGTACCCACATCCCTGCTATAAACCCAAAGAGAACTTCCA 1513  
501 ValGluAlaLysLeuProThrPheLysGlnAlaGlnGluLeuGluGln 517  
|||||  
1514 GTCGAGGCCAAGTGGCATGTTCAACAAGCTCAAGAGCTAGAGAAGG 1563  
517 yAlaAlaValSerGluGluProSerPheIleProGluAlaTrpSerAlaC 534  
|||||  
1564 AGCTGCTGTGTACAGAGAGGCGCTGTTTCATCCCCARAGGCGCTGGTCCGCT 1613  
534 yThrValThrCysGlyValGlyThrGlnValArgIleValArgCysGln 550  
|||||  
1614 GCACAGTCACCTGGGCTGTGGGACCCAGGTGGCAATAGTCAGTGGCAG 1663  
551 ValLeuLeuSerPheSerGlnSerValAlaAspLeuProLysLeuAspGluCy 567  
|||||  
1664 GTGCTCTGCTGTTCTCTCAGTCGCTGGCTGACCTGCCTATTGACGAGTG 1713  
567 sGluGlyProLysProAlaSerGlnArgAlaCysTyrAlaGlyProCysS 584  
|||||  
1714 TGAAGGCCCAAGCCAGCAGTCCCGAGCGCTGTTATGACAGGCCCATGCA 1763  
584 erGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheGly 600  
|||||  
1764 GCGSGGAAATTCCTGAGTCAACCCACAGACAGACAGATGGCTCTTTGGT 1813  
601 GlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluGlyPheTh 617  
|||||  
1814 GGCTCGAGGATTTGACGAGCTGTATGACTGGGAGTATAGGGGTTTCA 1863  
617 rLysCysSerCysGlyGlyValGlnGluAlaValValSerC 634  
|||||  
1864 CAAGTGTCCGAGTCTGTGGAGAGGTGCCAGAGGCTGTGTGAGCT 1913  
634 ysLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValThr 650  
|||||  
1914 GCTTGAACAACACAGACTCGGAGGCTGCTGAGGAGAACCTGTGGGTGACC 1963  
651 SerArgArgProProGlnLeuLeuLysSerCysAsnLeuAspProCysPr 667  
|||||  
1964 AGCCGCGGCGCCCAACAGCTCCTGAAGTCTCTGCAATTTGGATCCCTGCC 2013  
667 oAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysGlyV 684  
|||||  
2014 AGCAAGTGGGAAATGGCAAGTGGAGTCCATGTAGTCTACATGTGGGG 2063  
684 alGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArgGlu 700  
|||||  
2064 TCGGCTTACAGACCAGAGAGCTCTTCGACCCACCTGCTTCCAGAGAG 2113  
701 MetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLysPr 717  
|||||  
2114 ATGAATGAACAGTCACTCTGGCTGATGAGCTGTGCGCCAGGCCAAGCC 2163  
717 oSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTrp 734  
|||||  
2164 CAGCACGGTGCAGCTGTGAACCGCTTAAATGCCCCCAGCCCTGGTACC 2213  
734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750  
|||||  
2214 CTGCACAGTGGCAGCCGTTTCCAGACGTTGTGGCGGGGTGTTCAAGAA 2263

751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGluLe 767  
|||||  
2264 CGTGAGGTTCITTTGCAAGCAGCATGCTGATGSCAGCTTCTCTGGAGCT 2313  
767 uProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCysL 784  
|||||  
2314 TCCTGAGACCTTCTGTTCAAGTTCAAAACCTCGCTGCCAGCAAGCATGCA 2363  
784 ysLysAspAspCysProSerGluTrpLeuLeuSerAspTrpThrGluCys 800  
|||||  
2364 AGAAGATGACTGTCCAGCAGGTGGCTTCTCTCAGACTGGACAGAGTGT 2413  
801 SerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArgLy 817  
|||||  
2414 TCCACAAGCTCGGGGAAGGACCCAGACTCGAAGCGCATTTGCCGAAA 2463  
817 sMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysProp 834  
|||||  
2464 GATGCTGAAAACCGGCTCTCAACGGTTGTCAATTCACCCCTGTGCCGCG 2513  
834 roLeuProPheSerSerIleArgProCysMetLeuAlaThrCysAla 850  
|||||  
2514 CCCTGCGCTTCTCTCTCCATCAGGCCCTGTATGCTGGCAACCTGTGCA 2563  
851 ArgProGlyArgProSerThrLysHisSerProHisIleAlaAlaAlaAr 867  
|||||  
2564 AGGCCGCGGCGGCATCCACGAAGCACAGCCCGCACATCGCGGCGCCAG 2613  
867 gLysValTyrIleGlnThrArgArgGlnArgLysLeuHisPheValValG 884  
|||||  
2614 GAAGGCTTACATACAGACTCGCAGGCAGAGGAAGTGCACCTTCGTGTGG 2663  
884 lyGlyPheAlaTyrLeuLeuProLysThrAlaValValLeuArgCysPro 900  
|||||  
2664 GGGCTTCGCTTACCTGCTCCCCAAGACGGCGGTGGTCTGGCTGGCGCG 2713  
901 AlaArgArgValArgLysProLeuIleThrTrpGluLysAspGlyGlnHi 917  
|||||  
2714 GCGCGAGGGTCCGCAAGCCCTCATCACCTGGGAGAAGGACGGCCAGCA 2763  
917 sLeuIleSerSerThrHisValThrValAlaProPheGlyTyrLeuLysI 934  
|||||  
2764 CCTCATCAGCTCGACGCACGTACGGTGGCCCCCTTCGGCTATCTCAAGA 2813  
934 leHisArgLeuLysProSerAspAlaGlyValTyrThrCysSerAlaGly 950  
|||||  
2814 TCCACCGCTCAAGCCCTCGGATGCAGCGCTTACACCTGCTCAGCGGCG 2863  
951 ProAlaArgGluHisPheValIleLysLeuIleGlyGlyAsnArgLysLe 967  
|||||  
2864 CCGGCGCGGAGACACTTGTGATTAAGCTCATCGAGGCAACCGCAAGCT 2913  
967 uValAlaArgProLeuSerProArgSerGluGluValLeuAlaGlyA 984  
|||||  
2914 CGTGCGCGCGGCTTGAGCCCGCAAGTGAAGAGAGGTGCTTCGGGGGA 2963  
984 rGlyGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsnGly 1000  
|||||  
2964 GGAAGGGGGCGCCAGAGGCGCTGCAGACCCACCAACACAGAACGGG 3013  
1001 IlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlaAsnPr 1017  
|||||  
3014 ATCTTCTCAACGCGCAGCAAGCGGAGAAAGGGGCGCTGGCGCCCAAGCC 3063  
1017 oGlySerArgTyrAspAspLeuValSerArgLeuLeuGlnGlyGlyT 1034  
|||||  
3064 GGGAGCGGCTACGACGACCTCGTCTCCCGCTGCTGGAGCAGGCGGCT 3113  
1034 rpProGlyGluLeuAlaSerTrpGluAlaGlnAspSerAlaGluArg 1050  
|||||  
3114 GGCCCGGAGAGCTGCTGGCTTCGTGGGAGGCGCAGGACTCTGCGGAAAGG 3163  
1051 AsnThrThrSerGluGluAspProGlyAlaGluGlnValLeuLeuHisIle 1067

||||| 3164 AACACGACCTCGGAGAGGAGCCGGGTGCAGAGCAAGTGCTCTCCACCT 3213  
1067 uProPheThrMetValThrGluGlnArgArgLeuAspAspIleLeuGlyA 1084  
||||| 3214 GCCCTTCACCATGGTGACCGAGCAGCGCGCTCGGACACATCTGGGGA 3263  
1084 snLeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeu 1100  
3264 ACCTTCGCCAGCAGCCGAGAGTGCGCGACCTCTACAGCAAGCACCTG 3313  
1101 ValAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAs 1117  
||||| 3314 GTGCCCGAGCTGGCCCGAGGAGATCTCCGACAGCCACCTGGAGCACCAGGA 3363  
1117 pThrLeuLeuLysProSerGluArgArgThrSerProValThrLeuSerP 1134  
3364 CACGCTCCTGAAGCCCTCGGAGCGCAGGACTTCCCCAGTGACTCTCTCGC 3413  
1134 roHisLysHisValSerGlyPheSerSerSerLeuArgThrSerSerThr 1150  
3414 CTCATAAACACGTGTCGTGGTTCAGCAGCTCCCTGCGGACCTCTCCACC 3463  
1151 GlyAspAlaGlyGlySerArgArgProHisArgLysProThrIleLe 1167  
3464 GGGACCCCGGGGAGGCTCTCGAAGGCCACACCGCAAGCCACCATCCT 3513  
1167 uArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrH 1184  
3514 GCGCAAGATCTACGCGCCCGACAGCTCTCAGCCTCGGAGGTGTCACCC 3563  
1184 lsLeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeu 1200  
3564 ACCTGGGGCAGACGGTGGCCCTGGCCAGCGGACACTCAGTGTTCCTCTG 3613  
1201 HisCysGluAlaIleGlyHisProArgProThrIleSerTrpAlaArgAs 1217  
3614 CACTGTGAGGCCATCGGCCACCCACCGCCCTACCATCAGCTGGGCCAGGA 3663  
1217 nGlyGluGluValGlnPheSerAspArgIleLeuLeuGlnProAspAsps 1234  
3664 TGGAGAAGAGTTCAGTTCAGTGACAGGATCTTCTACAGCCAGATGATT 3713  
1234 erLeuGlnIleLeuAlaProValGluAlaAspValGlyPheTyrThrCys 1250  
3714 CCTTACAGATCTTGGCCACCATGTGAACAGATGTGGGTTCCTACACTGC 3763  
1251 AsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValThrLe 1267  
3764 AATGCCACCAATGCTTGGATAGGACTCTGTCTCOATTGCCGTACATT 3813  
1267 uAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThrG 1284  
3814 AGCAGGAAGCCACTAGTGAAGAGTGCACCAATGACAGTGATCAACACGG 3863  
1284 LuLysProAlaValThrValAspIleGlySerThrIleLysThrValGln 1300  
3864 AGAAGCTTGGAGTCCAGTCCATATAGGAAGCACCATCAAAACAGTGCAG 3913  
1301 GlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317  
3914 GGAGTGAATGTGACAATCACTAGTGCAGGTTGCAGGAGTGCTGAAGCTGA 3963  
1317 uValThrTrpPheArgAsnLysSerLysLeuGlySerProHisHisLeuH 1334  
3964 AGTCACCTGTGTTACAGGAATAAAAGCAAACTGGGCTCCCGCACCACCTGC 4013  
1334 isGluGlySerLeuLeuThrAsnValSerSerSerAspGlnGlyLeu 1350  
4014 ACGAAGGCTCCTGCTCCTCACAACCGTGTCTCCTCGGATCAGGGCTG 4063  
1351 TyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGluSerThrGl 1367  
|||||

4064 TACTCTCTGCAGGGGGCCCAATCTTCATGGAGAGCTGACTGAGACACCCA 4113  
1367 nLeuLeuIleLeuAspProProGlnValProThrGlnLeuGluAspIleA 1384  
||||| 4114 GCTGCTGATCCTAGATCCCCCCCAGTCCCCACACAGTTGGAAGACATCA 4163  
1384 rgAlaLeuLeuAlaAlaThrGlyProAsnLeuProSerValLeuThrSer 1400  
4164 GGGCCTTGCTGCTGCCACTGGACCGAACCTTCCCTCAGTGTGACGTCT 4213  
1401 ProLeuGlyThrGlnLeuValIleuAspProGlyAsnSerAlaLeuLeuGl 1417  
4214 CCTCTGGGAACACAGCTGGTCTGGATCCTGGGAATCTGCTCTCCTTGG 4263  
1417 yCysProIleLysGlyHisProValProAsnIleThrTrpPheHisGlyG 1434  
4264 CTGCCCCATCAAGGTCAACCTGTCCCTTAATATACCTGGTTTCATGGTG 4313  
1434 lyGlnProIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaAla 1450  
4314 GTACGCCAATTGTCACCTGCCACAGGACTGACGCATCACATCTTGGCAGCT 4363  
1451 GlyGlnIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyGluPh 1467  
4364 GGACAGATCCTTCAAGTTGCAACCTTAGCGGTGGGTCTCAAGGGGAATT 4413  
1467 eSerCysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerL 1484  
4414 CAGTGCCTTGTCTCAGAAATGAGGAGGGTGTCTCATGCAAGAGGCATCTT 4463  
1484 euValIleGlnAspTyrTrpSerValAspArgLeuAlaThrCysSer 1500  
4464 TAGTGATCCAAGATTACTGGTGGTGTGTGGACAGACTGGCAACCTGTCTCA 4513  
1501 AlaSerCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLe 1517  
4514 GCCCTCTGTGTAACTGGGGGGTTTACAGAGCCCGCTTGSAGGTGGCTGCT 4563  
1517 uAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProA 1534  
4564 GAACAGCAGCGAGGTCAACCTTGCCACTGCGCAGGGAAGGTTCGCCCTG 4613  
1534 laValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMet 1550  
4614 CGGTGCAGCCCATCGGTGCAACCGGAGAGACTGCCCTTCTCGGTGGATG 4663  
1551 ValThrSerTrpSerAlaCysThrArgSerCysGlyGlyValGlnTh 1567  
4664 GTGACCTCTGCTGCTGCTGTACCCGGAGCTGTGGGGAGGTGTCCAGAC 4713  
1567 rArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProv 1584  
4714 CCGCAGGTGACCTGTCAAAAGCTGAAAGCTCTGGGATCTCCACCCCTG 4763  
1584 alSerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGln 1600  
4764 TGTCCAATGACATGTGCACCCAGGTGCGCAAGCGGCTGTGGACACCCAG 4813  
1601 AlaCysAsnGlnGlnLeuCysValGluTrpAlaPheSerSerTrpGlyGl 1617  
4814 GCCTGTAAACAGCAGCTGTGTGTGAGGTGGCTTCTCCAGCTGGGGCA 4863  
1617 nCysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnV 1634  
4864 GTGCAATGGGCTTGCATCGGGCTTCACTAGCTGTGCAACACACACAAG 4913  
1634 alPheCysGlnThrArgAspGlyIleThrLeuProSerGluGlyCysSer 1650  
4914 TCTTCTGCCACACAGCGGATGGCATCCTTACCATCAGAGCAGTGCAGT 4963  
1651 AlaLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysSe 1667  
4964 GCTCTCCGAGACCTGTGAGCACCCAGCAACTGCTGTCAGAGGCTGCAG 5013



1667 rValHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyA 1684  
 |||||  
 5014 TGTACACTGGAGAGTCAGCTGTGGACCTGTGCACAGCTACTGTGGCA 5063  
 |||||  
 1684 sNtYrGlyPheGlnSerArgValGluCysValHisAlaArgThrAsn 1700  
 |||||  
 5064 ACTACGGCTTCAGTCCCGCGGTGGAGTGTGTGCATGCCCGCACCAAC 5113  
 |||||  
 1701 LysAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTr 1717  
 |||||  
 5114 AAGGCAGTGCCTGAGCACCTGTCTCTGGGGCCCGCGCTGCCAACYG 5163  
 |||||  
 1717 pGlnArgCysAsnIleThrProCysGluAsnMetGluCysArgAspThrT 1734  
 |||||  
 5164 GCAGCGCTGCACATCACCCCATGTGAAACATGGAGTGCAGAGACCA 5213  
 |||||  
 1734 hrArgTrpCysGluLysValLysGlnLeuLysCysGlnLeuSerGln 1750  
 |||||  
 5214 CCAGGTACTGCCGAGAAGGTGAACAGCTGAAACTCTGCCAACCTCAGCCAG 5263  
 |||||  
 1751 PheLysSerArgCysCysGlyThrCysGlyLysAla 1762  
 |||||  
 5264 TTTAAATCTCGTGTGCTGGAATCTGTGCAAGCG 5299  
 |||||

seq\_name: /SIDS1/gcgdata/hold-geneseg/geneseqn-emb1/NA2001A.DAT:AAF97926

seq\_documentation\_block:

ID AAF97926 standard; cDNA: 3559 BP.

XX AC AAF97926;

XX DT 01-JUN-2001 (first entry)

XX DE Human secreted protein cDNA, SEQ ID NO: 53.

XX KW Human; secreted protein; immunomodulatory; antisclerotic;  
 KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiac;  
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 KW neotropic; anticonvulsant; anti-Alzheimer's; antiparkinsonian;  
 KW antimicrobial; vulnery; vaccine; gene therapy; cancer;  
 KW protein coordinate data; infection; ss.

XX OS Homo sapiens.

XX FN WO200121658-A1.

XX PD 29-MAR-2001.

XX PF 22-SEP-2000; 2000WO-US26013.

XX PR 24-SEP-1999; 99US-0155709.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;

XX DR WPI; 2001-235311/24.

XX PT Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -

XX PS Claim 1; Page 747-748; 890pp; English.

XX CC The present sequence encodes one of 32 novel human secreted polypeptides.  
 CC The nucleic acid molecules and polypeptides they encode may be used in  
 CC the prevention, diagnosis and treatment of diseases such as  
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus  
 CC and human immuno-deficiency virus (HIV) infections), hyperproliferative  
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases

CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the  
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid  
 CC sequences in samples. The polypeptides may be used as antigens in the  
 CC production of antibodies and in assays to identify modulators of  
 CC their expression and activity.

XX SQ Sequence 3559 BP; 807 A; 1037 C; 1005 G; 698 T; 12 other;

alignment\_scores:

Quality: 6176.50 Length: 1201  
 Ratio: 5.279 Gaps: 3  
 Percent Similarity: 97.419 Percent Identity: 96.919

alignment\_block:

US-10-044-807-2 x AAF97926

Align seg 1/1 to: AAF97926 from: 1 to: 3559

73 TyrArgThrCysSerAsnValAspCysProGluAlaGlyAspPheAr 89  
 :: |||||  
 13 TTTTAAACATGCAGCCCTTTGGAATGCCCAACCAAGCMSTGATTTCCG 62  
 |||||  
 89 gAlaGlnGlnCysSerAlaHisAsnValLysHisGlyGlnPheT 106  
 |||||  
 63 AGCTCAGCAATGCTCAGCTCATATGATGCAAGCACCAGTGGCAGTTTT 112  
 |||||  
 106 yrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeuLys 122  
 |||||  
 113 ATGAATGGCTTCTGTGTCTTAATGACCTGACCAACCCATGTTCACTCAAG 162  
 |||||  
 123 CysGlnAlaLysGlyThrLeuValValGluLeuAlaProLysValle 139  
 |||||  
 163 TGCCAAAGCCAAAGGAACACCCCTGTTGTGTAAGTACAGCCTAAGTCTT 212  
 |||||  
 139 uAspGlyThrArgCysTrpThrGluSerLeuAspMetCysIleSerGlyL 156  
 |||||  
 213 AGATGGTACGGTGTGCTATACAGAAATCTTTGGATATCTGCATCAGTGGT 262  
 |||||  
 156 euCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLysGlu 172  
 |||||  
 263 TATGCCAAATTTGTTGGTGGATCACCAGCTGGGAAGCACCCTCAAGGAA 312  
 |||||  
 173 AspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuValAr 189  
 |||||  
 313 GATAACTGTGGGTCTGCAACGGAGATGGTCCACCTGCCGGCTGGTCCG 362  
 |||||  
 189 gGlyGlnTrpLysSerGlnLeuSerAlaThrLysSerAspThrValV 206  
 |||||  
 363 AGGGCAGTATAATCCAGCTCTCCGCAACCAAAATCGGATGATGACTGG 412  
 |||||  
 206 aAlaIleProTrpGlySerArgHisIleArgLeuValLysGlyPro 222  
 |||||  
 413 TTGCAATTCCTATGGAAGTAGACATATTCGCCCTTGTCTTAAAGGTCCT 462  
 |||||  
 223 AspHisLeuTrpLeuGluThrLysThrLeuGlnGlyThrLysGlyGluAs 239  
 |||||  
 463 GATCACTTATATCTGGAACCAAAACCCCTCCAGGGGACTAAAGGTCAAAA 512  
 |||||  
 239 nSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerValAsp 256  
 |||||  
 513 CAGTCTCAGCTCCACAGGAACCTTCTCTGTCGACAAATTTAGTGTGCACT 562  
 |||||  
 256 heGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeuThr 272  
 |||||  
 563 TCCAGAAATTTCCAGAACAAAGAGATACTAGAAATGGCTGGACCACTACA 612  
 |||||

273 AlaAspPheIleValIysIleArgAsnSerGlySerAlaAspSerThrVa 289  
|||||GCAGATTTTCATGTCAGATTCTGAATTCGGGCTCCGCTGCACAGTACAGT 662  
289 IGlInPheIlePheTyrGlnProIleIleHisArgTrpArgGluThrAspP 306  
|||||CCAGTTCATCTTCATCAACCCCATCATCCACCCGATGCCGATGAGGAGACGATT 712  
306 hePheProCysSerAlaThrCysGlyGlyTyrGlnLeuThrSerAla 322  
|||||TCCTTCTGCTCAGCAACTCTGGAGGAGGTATCAGTCGACATCGGCT 762  
323 GluCysTyrAspLeuArgSerAsnArgValAlaAspGlnTyrCysHI 339  
|||||GAGTGTACGATCTGAGGAGCAACCGTGTGTGCTGACCAATCTGTCA 812  
339 sTyrTyrProGluAsnIleLysProLysProLysLeuGlnGluCysAsnL 356  
813 CTATTACCCAGAGAAATCAAAACCCAAACCCAAAGCTTCAGGAGTGCAACT 862  
356 euAspProCysProAlaSerAspGlyTyrIysGlnIleMetProTyrAsp 372  
863 TGGATCTTGTCCAGCC..... 879  
373 LeuTyrHisProLeuProArgTrpGluAlaThrProTrpThrAlaCysSe 389  
880 .....AGTGGGAGGCCACCCCATGGACCGCGTGCTC 911  
389 rSerSerCysGlyGlyClyIleGlnSerArgAlaValSerCysValGluG 406  
912 CTCCTCGTGGGGGGGGATCCAGAGCGGGCAGTTTCTGTGTGGAGG 961  
406 luAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMetTyr 422  
962 AGGACATCCAGGGGCATGTCACCTTCAGTGGAGAGTGGAATGCATGTAC 1011  
423 ThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysProLy 439  
1012 ACCCCTAAGATGCCCATCGCGCAGCCCTGCACATTTTGTGACCTGCGCTAA 1061  
439 sTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGluGlyL 456  
1062 ATGGCTGGCACAGAGTGTCTCCGTGCACAGTGACATGGCCAGGGCC 1111  
456 euArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThrGly 472  
1112 TCAGATACCGTGTGTCTCTGCATCGACCATCGAGGAATGCACACAGGA 1161  
473 GlyCysSerProLysThrLysProHisIleLysGluGluCysIleValPr 489  
1162 GGCTGTAGCCCAAAACAAAGCCCCACATATAAAGAGGAATGCATCGTACC 1211  
489 oThrProCysTyrLysProLysGluLysLeuProValGluAlaLysLeup 506  
1212 CACTCTCTGCTATTAACCCAAAGAGAAACTTCCAGTCGAGGCCAAGTTGC 1261  
506 roTrpPheLysGlnAlaGlnGluLeuGluGluGlyAlaAlaValSerGlu 522  
1262 CATGGTTCAACAAGCTCAAGAGCTAGAGAAGGAGCTGCTGTGTCCAGAG 1311  
523 GluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCysGI 539  
1312 GAGCCCTCGTTCATYTCARAGGCTGTGGCCCTGCACATCACCTGTGTGG 1361  
539 yValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerPheS 556  
1362 TGGGGGACCCARGTCGAATAGTACGTAGTGCCAGGTGCTCTCTCTTCT 1411  
556 erGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLysPro 572  
1412 CTCAGTCCGTGGTGCCTGCTATTGACGAGTGTGAAGGGCCCAAGCCA 1461  
573 AlaSerGlnArgAlaCysTyrAlaGlyProCysSerClyGluIleProGI 589

1462 GCATCCAGCGTGCCTGTATGCAGGCCCATGCAGCGGGGAAATTCCTGA 1511  
589 uPheAsnProAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheA 606  
1512 GTTCAACCCAGACAGACATGGGCTTTTGGTGGCTGCAGGATTTCCG 1561  
606 spGluLeuTyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSer 622  
1562 ACAGCTGTATGACTGGAGTATGAGGGGTTCACCAAGTGTCTCGAGTCC 1611  
623 CysGlyGlyGlyValGlnGluAlaValIserCysLeuAsnIysGlnTh 639  
1612 TGTGGAGGAGTGTCCAGGAGGCTGTGCTGAGCTGCTTGAACAAACAGAC 1661  
639 rArgGluProAlaGluAsnLeuCysValThrSerArgArgProProG 656  
1662 TCGGAGCGCTGCTGAGAGAACCTGTGCTGACAGCGCGCGGCCCCAC 1711  
656 InLeuLeuLysSerCysAsnLeuAspProCysProAlaArgTrpGluIle 672  
1712 AGCTCCTGAAGTCTCTCAATTTGATCCCTGCCAGCAAGGTGGGAATT 1761  
673 GlyLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnThrAr 689  
1762 GGCAAGTGGAGTCCATGTAGTCTCACATGTGGGTGGCTTACAGACCAG 1811  
689 qAsnValPheCysSerHisLeuLeuSerArgGluMetAsnGluThrValI 706  
1812 AGAGCTCTTTCGACGACCCTGTCTTCCAGAGATGAATGAACAGTCA 1861  
706 leLeuAlaAspGluLeuCysArgGlnProLysProSerThrValGlnAla 722  
1862 TCCTGGCTGATGAGTGTGTGCCAGCCCAAGCCAGCAGCGTCAAGCT 1911  
723 CysAsnArgPheAsnCysProProAlaTrpTyrProAlaGlnTrpGlnPr 739  
1912 TGTAAACCGCTTAAATTCGCCCCCGACCTGTGTACCTGCACAGTGGCAGCC 1961  
739 oCysSerArgThrCysGlyGlyValGlnLysArgGluValLeuCysL 756  
1962 GTGTTCAGAACGCTGTGGCGGGGTTCAGAAACGCTGAGGTTCTTTGCA 2011  
756 ysGlnArgMetAlaAspGlySerPheLeuGluLeuProGluThrPheCys 772  
2012 AGCAGCGCATGGCTGATGGCAGCTTCTCGAGCTTCTGAGACCTTCTGT 2061  
773 SerAlaSerLysProAlaCysGlnAlaCysLysLysAspCysPr 789  
2062 TCAGCTTCAAAACCTGCTGCCAGCAAGCATGCAAGAAAGATGACTGTCC 2111  
789 oSerGluTrpLeuLeuSerAspTrpThrGluCysSerThrSerCysGlyG 806  
2112 CAGGAGTGGCTTCTTCACAGCTGGACAGAGTGTTCACAGCTGCGGGG 2161  
806 luGlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThrGly 822  
2162 AAGSCACCCAGACTCGAAGCGCCATTTGCCAAAGATGCTGAAACCGGC 2211  
823 LeuSerThrValValAsnSerThrLeuCysProProLeuProPheSerSe 839  
2212 CTCTCAACGGTTGTCAATTCACCTGTGCCGCCCTGCTTCTCTTC 2261  
839 rSerIleArgProCysMetLeuAlaThrCysAlaArgProGlyArgProS 856  
2262 CTCCATCAGCCCTGTATGCTGGCAACTGTGCAAGCCCGCGGCGGCAT 2311  
856 erThrLysHisSerProHisIleAlaAlaAlaArgLysValTyrIleGln 872  
2312 CCAGGAAGCAGACCGCCACATCGCGCGCCAGGAAGGTCTACATACAG 2361  
873 ThrArgArgGlnArgLysLeuHisPheValValGlyGlyPheAlaTyrLe 889  
|||||

```
2362 ACTCGCAGGCAGAGGAAGCTGCTCGTGGTGGGGGCTTCGCCTACCT 2411
889 uLeuProLysThrAlaValValLeuArgCysProAlaArgValArgL 906
|||||
2412 GCTCCCAAGACGCGGTGTGTGCGCTGCCGCGCGAGGGTCGCA 2461
906 ysProLeuIleThrTrpGluLysAspGlyGlnHisLeuIleSerSerThr 922
|||||
2462 AGGCCCTCATCAGCTGGGAGAGGACGGCCAGCAGCACTCATCAGCTCAGC 2511
923 HisValThrValAlaProPheGlyTyrLeuLysIleHisLeuLysPr 939
|||||
2512 CAGTCAGGTGGCCCTTCCTGGCTATCTCAAGATCCACCGCTCAAGCC 2561
939 oSerAspAlaGlyValTyrThrCysSerAlaGlyProAlaArgGluHisP 956
|||||
2562 CTCGGATGCAGGCGTCTACACCTGCTCAGCGGGCCGCGGAGCACT 2611
956 heValIleLysLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeu 972
|||||
2612 TTGTGATTAAAGCTCATCGGAGGCAACCGCAAGCTCGTGGCCGCGCCTTG 2661
973 SerProArgSerGluGluGluValLeuAlaGlyValArgLysGlyProLy 989
|||||
2662 AGCCCGAAGACTGAGGAGAGGTGCTTCGCGGAGGAGGCGGCCCGAA 2711
989 sGluAlaLeuGlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlyS 1006
|||||
2712 GGAGGCCCTGCAGACCCACAAACACCAAGACGGGATCTTCTCCAACGCA 2761
1006 erLysAlaGluLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAsp 1022
|||||
2762 GCAAGCGGAGAGCGGGGCTGGCGCCCAACCGGGGAGCGCTACGAC 2811
1023 AspleuValSerArgLeuLeuGluGlnGlyTyrProGlyGluLeuLe 1039
|||||
2812 GACTCGTCTCCGCGCTGCTGGACAGCGGGCTGGCCCGGAGAGCTGCT 2861
1039 uAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrserGluG 1056
|||||
2862 GGCCTCGTGGGAGCGCAGGACTCTGCGGAAAGAACACGACCTCGGAGG 2911
1056 luAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMetVal 1072
|||||
2912 AGGACCGGGTGCAGAGCAAGTGTCTGCACTGCGCTTCCACCATGGTG 2961
1073 ThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGlnGlnPr 1089
|||||
2962 ACCGAGCAGCGCGCTGGACGACATCTCTGGGNAACCTCTCCACGACGCC 3011
1089 oGluGluLeuArgAspLeuTyrSerLysHisLeuValAlaGlnLeuAlaG 1106
|||||
3012 CGAGGAGCTGCGGACCTCTACAGCAAGCACCTGTTGGCCAGCTGGGCC 3061
1106 lnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeuLysPro 1122
|||||
3062 AGGAGATCTTCCGACGACCTGAGCACCAGGACAGCTCTCTGTAAGCCC 3111
1123 SerGluArgArgThrSerProValThrLeuSerProHisLysHisValse 1139
|||||
3112 TCGGAGCGCAGGACTTCCCGAGTCACTCTCTCGCTCATAAACACGTGTC 3161
1139 rGlyPheSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGlyG 1156
|||||
3162 TGGCTTTCAGAGCTCCCTGGGAGCTCTCCACCGGGGAGCGCGGGGAG 3211
1156 lySerArgArgProHisArgLysProThrIleLeuArgLysIleSerAla 1172
|||||
3212 GCTCTGAAGGCCACACCGAAGCCCACTCTCTGCGCAAGATCTCAGCG 3261
1173 AlaGlnGlnLeuSerAlaSerGluValValThrHisLeuGlyGlnThrVa 1189
|||||
3262 GCCCAGGACTCTCAGCTCGGAGGTGGTCAACCCACTGGGGGACAGCGGT 3311
```

```
1189 lAlaLeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluAlaIleG 1206
|||||
3312 GGCCTGGCCAGCGGACACTGAGTGT.CTCTGCACTGTGAGGCCATCG 3360
1206 lyHisProArgProThrIleSerTrpAlaArgAsnGlyGluGluValGln 1222
|||||
3361 GCNACCCAAAGCCCTACCATCAGCTGGCGCAGGAATGGAGAAGAAGT.CAG 3409
1223 PheSerAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeu.A 1239
|||||
3410 TTCAGTCACAGGATCTCTACAGCAGATGATTCYTTACAGATCTTGGG 3459
1239 laProValGlu.AlaAspValGlyPheTyrThrCysAsnAlaThrAsnAl 1255
|||||
3460 CACCAGTGGAAAGCAGATGGGTTTCTACACTTGCATGGGCAC.AATGC 3508
1255 aLeuGlyTyrAspSerValSerIleAlaValThrLeuAlaGlyLysPro 1271
|||||
3509 CTTGGGATACGAMTTTGTCTCCATTGGCGTCACATTACAGAAAAGCCC 3557
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.AAA47482

seq\_documentation\_block:

ID AAA47482 standard; cDNA; 2689 BP.

XX AAA47482;

XX AC

XX DT 20-OCT-2000 (first entry)

XX XX Human TANGO 224 coding sequence (form 1).

XX TANGO: 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
KW inflammatory bowel disease; septic shock; ulcerative colitis;  
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
KW systemic lupus erythematosus; transgenic animal; diagnosis;  
KW prognosis; prophylactic; therapeutic; human; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 67..2688

XX FT /\*tag= a

XX FT /product= TANGO 224

XX XX

XX PN WO200039284-A1.

XX PD 06-JUL-2000.

XX XX 23-DEC-1999; 99WO-US31025.

XX PF 30-DEC-1998; 98US-0223546.

XX XX (MILL-) MILLENNIUM PHARM INC.

XX XX Holtzman DA;

XX XX WPI; 2000-465743/40.

XX DR P-PSDB; AA901431.

XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases  
XX Claim 1; Fig 30; 209pp; English.

XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
CC agents for regulating cellular processes like asthma, graft  
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,

CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
 CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
 CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
 CC lupus erythematosus. The nucleic acids are also useful for producing  
 CC transgenic animals and the TANGO polypeptides themselves. Partial  
 CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
 CC forensic biology, for diagnostic assays, prognostic assays,  
 CC pharmacogenomics and for monitoring clinical trials. TANGO  
 CC polypeptides are suitable for both prophylactic and therapeutic  
 CC methods for treating a subject at risk of a disorder or having a  
 CC disorder associated with aberrant TANGO expression. A wide range  
 CC of cellular disorders can be treated.  
 XX

SQ Sequence 2689 BP: 633 A; 739 C; 734 G; 583 T; 0 other;

alignment\_scores:  
 Quality: 4797.50 Length: 874  
 Ratio: 5.546 Gaps: 1  
 Percent Similarity: 98.970 Percent Identity: 98.856

alignment\_block:

US-10-044-807-2 x AAA47482 ..

Align seg 1/1 to: AAA47482 from: 1 to: 2689

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17  
 67 ATGAATGCTGCCGTGGGCAACTCCTGGCACACTGCTCTCTTTCTGGC 116  
 17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluAlaSPArgasp 34  
 117 TTTCTCTCTCTCAGTTCAGAGACCGCACGCTCCGAGGAGCGGGAGC 166  
 34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50  
 167 GCCTGGGATGCTGGGGCCCATGGAGTGAATGCTCACGACCTGGCGG 216  
 51 GlyClyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG 67  
 217 GGAGGGCCCTCCACTCTCTGAGGCGGTGCTGAGCAGCAAGAGCTGTGA 266  
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84  
 267 AGGAAGAAATATCCGATACAGACATGCAGTAATGTGGACTGCCACAG 316  
 84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100  
 317 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 366  
 101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117  
 367 CACCATGGCCAGTTTATGAATGGCTTCCTCTGTCTAATGACCTGACAA 416  
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134  
 417 CCCATGTCTCACTCAAGTGGCCAAAGCAACCCCTGGTGTGTTGAAC 466  
 134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150  
 467 TAGCACCCTAAGGCTTAGATGGTACGCTGTGCTATACAGAACTTTGGAT 516  
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG 167  
 517 ATGTGCATCAGTGGTTATGCCAAATTTGTTGGCTGGCATCACCAGCTGGG 566  
 167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184  
 567 AAGCACCCTCAAGGAAGATTAACGTGGGGTCTGCACCGAGATGGGTCCA 616  
 184 hrCysArgLeuValArgGlyClnTyrLysSerGlnLeuSerAlaThrLys 200  
 617 CCTGCCGCTGGTCCGAGGCGAGTATAAATCCCGAGCTCTCCGCAACCAAA 666

201 SerAspSPThrValValAlaIleProTyrGlySerArgHisIleArgLe 217  
 667 TCGCATCATCTGGTGGTGCATTTCCCTATGGAAGTAGACATATTCGGCT 716  
 217 uValLeuLysGlyProaspHisLeuTyrLeuGluThrLysThrLeuGlnG 234  
 717 TGTCTTAAAGGTCTCTGATCACTTATATCTGAAACCAAAACCTCCAGG 766  
 234 lyThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250  
 767 GGACTAAGGTGAACACAGTCTCAGCTCCACAGGAACITTCCTTGTGGAC 816  
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267  
 817 AATTCTAGTGGACTTCCAGAAATTTCCACACAAAGAGATACTGAGAAT 866  
 267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284  
 867 GGCTGGACCACTACAGCAGATTTTCAATGTCAAGATTGTAACCTCGGCT 916  
 284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300  
 917 CCGCTGACAGTACAGTCCAGTTCATCTTATCAACCCCATCATCCACGA 966  
 301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTY 317  
 967 TGGAGGAGACGGATTTCTTTCTTCTGCTCAGCAACCTGTGGAGGAGTTA 1016  
 317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334  
 1017 TCAGCTGACATCGGCTGAGTCTAGATCTGAGGAGCAACCGCTGTGGTTG 1066  
 334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350  
 1067 CTGACCAATACTGTCACTATTACCCAGAGAACATCAAAACCCCAACCC 1116  
 351 LeuGlnGluCysAsnLeuAspProCysProAla ..... 361  
 1117 CTTCAGGAGTGCAACTTGGATCTTGTCCAGCCAGGAGGTGGCATTAT 1166  
 362 .....SerAspGlyTyrLysGlnIleMetProTyrAspLeuTyrH 375  
 1167 TGTCTCTAACAGTGACGGATACAGCAGATCATGCCCTTATGACCTTACC 1216  
 375 isProLeuProArgTrpGluAlaThrProTrpThrAlaCysSerSerSer 391  
 1217 ATCCCTTCTCTCGTGGGAGCCACCCCATGGACCGGTGCTCTCTCTCG 1266  
 392 CysGlyGlyGlyIleGlnSerArgAlaValSerCysValGluGluAspI 408  
 1267 TGTGGGGGGGCATCCAGAGCGGGCAGTTTCTGTGTGGAGGAGACAT 1316  
 408 eGlnGlyHisValThrSerValGluGluTrpLysCysMetTyrThrProL 425  
 1317 CCAGGGGCATGTCACTTCAGTGGAGAGTGGAAATGATGTACACCCCTA 1366  
 425 ysMetProIleAlaGlnProCysAsnIlePheAspCysProLysTrpLeu 441  
 1367 AGATGCCCATCGCGCAGCCCTGCAACATTTTGTACTGCCCTAATATGGCTG 1416  
 442 AlaGlnGluTrpSerProCysThrValThrCysGlyGlnGlyLeuArgTy 458  
 1417 GCACAGGAGTGGTCTCGGTGCACAGTGCACATGTGGCCAGGCGCTCAGATA 1466  
 458 rArgValValLeuCysIleAspHisArgGlyMetHisThrGlyCys 475  
 1467 CCGTGTGGTCTCTTGCATCGACCATCGAGGAATGCACACAGGAGGCTGTA 1516  
 475 erProLysThrLysProHisIleLysGluGluCysIleValProThrPro 491  
 1517 GCCCAAAACAAAGCCCAACATAAAGAGGAATGCATCGTACCCACTCCC 1566

492 CysTyrLysProLysGluLysLeuProValGluAlaLysLeuProTrpPh 508  
 1567 TGGTATAAAACCCAAAGAGAACTTCCAGTCGAGGCCAAAGTTCATGGTT 1616  
 508 eLysGlnAlaGlnGluLeuGluGlyAlaAlaValSerGluLysProS 525  
 1617 CAAACAAAGCTCAAGAGCTAGAACAGAGAGCTGCTGTCTCAGAGAGCCCT 1666  
 525 erPheIleProGluAlaTrpSerAlaCysThrValThrCysGlyValGly 541  
 1667 CGTTCATCCAGAGGCTGTGTCGCCCTGCACAGTCACCTGTGTGTGGG 1716  
 542 ThrGlnValArgIleValArgCysGlnValLeuLeuSerPheSerGlnSe 558  
 1717 ACCCAGGTGGAAATAGTCAGGTGCCAGGTCTCTCTCTCTCTCAGTC 1766  
 558 rValAlaAspLeuProIleAspGluCysGluGlyProLysProAlaSerG 575  
 1767 CHTGGCTGACCTGCCATTATGACAGAGTGTGAAGGCCCAACCCAGCATCCC 1816  
 575 InArgAlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsn 591  
 1817 AGCGTGCTCTTATGACAGGCCCATGACGGGGGAAATTCCTGAGTTCAAC 1866  
 592 ProAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheAspGluLe 608  
 1867 CCAGACGACAGATGGCTCTTTGGTGGCTGCAGGATTTCGACGAGCT 1916  
 608 uTyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG 625  
 1917 GATGACTGGAGATGAGGGGTTCACCAAGTCTCGAGTCTCTGTGGAG 1966  
 625 lyGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGlu 641  
 1967 GAGGTGTCCAGGAGGCTGTGTGAGCTGTGTAACAAACAGACTCGGGAG 2016  
 642 ProAlaGluLysLeuCysValThrSerArgArgProGlnLeuLe 658  
 2017 CTTGCTGAGAGAACCTGTGCTGAGCAGCGCGCGCCGCCACAGCTCCT 2066  
 658 uLysSerCysAsnLeuAspProCysProAlaArgTrpGluIleGlyLys 675  
 2067 GAAGTCTGCAATTTGGATCCCTGCCAGCAAGTGGGAAATTTGCAAGT 2116  
 675 rpSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspVal 691  
 2117 GGAGTCCATGTAGTCTCATGTGTGGGTCTGCTTACAGACAGAGACGTC 2166  
 692 PheCysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAl 708  
 2167 TTTCTGAGCCACCTGCTTTCCAGAGAGATGAATGAACAGTCATCTCTGGC 2216  
 708 aAspGluLeuCysArgGlnProLysProSerThrValGlnAlaCysAsnA 725  
 2217 TGATGAGCTGTGTGCCAGCCCAAGCCACGACGTCGAAGCTTGTAAAC 2266  
 725 rgPheAsnCysProProAlaTrpTyrProAlaGlnTrpGlnProCysSer 741  
 2267 GCTTTAATTGCCCGCCAGCTGTGTACCTGTGCACAGTGGCAGCGCTGTCC 2316  
 742 ArgThrCysGlyGlyValGlnLysArgGluValLeuLeuCysLysGlnAr 758  
 2317 AGAACGTGTGGCGGGGTGTTCAGAACGCTGAGGTCTTTTGAAGACAGCG 2366  
 758 qMetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlas 775  
 2367 CATGGCTGATGGCAGCTTCTGTGAGCTTCTGTGACCTTCTGTTCAGCTT 2416  
 775 erLysProAlaCysGlnGlnAlaCysLysLysAspCysProSerGlu 791  
 2417 CAAACCTGCTGCCAGCAAGCATGCAAGAAAGATGACTGTCTCCAGCGAG 2466  
 792 TrpLeuLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyTh 808

2467 TGGCTTCTCTCAGACTGGACAGAGTGTTCACAAAGCTGCGGGGAGGCAC 2516  
 808 rGlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerT 825  
 2517 CCAGACTCGAAGCGCCATTTGCCGAAAGATGCTGAAAAACCGGCATCTCAA 2566  
 825 hrValValAsnSerThrLeuCysProProLeuProPheSerSerSerIle 841  
 2567 CGGTGTCAATTCACCTGTGCCCGCCCTGCGCTTCTCTCTCTCTCCATC 2616  
 842 ArgProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLys 858  
 2617 AGGCCCTGTATGCTGGCAACCTGTGCAAGGCCGCGCGGCATCCACGAA 2666  
 858 shISerProHisIleAlaAla 865  
 2667 GCACAGCCCCGCACATCGGGCC 2688  
 seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAA47458  
 seq\_documentation\_block:  
 ID AAA47458 standard; cdna; 2689 BP.  
 AC AAA47458;  
 XX  
 DT 20-OCT-2000 (first entry)  
 XX  
 DE Human TANGO 224 coding sequence.  
 XX  
 KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
 KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
 KW inflammatory bowel disease; septic shock; ulcerative colitis;  
 KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
 KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
 KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
 KW systemic lupus erythematosus; transgenic animal; diagnosis;  
 KW prognosis; prophylactic; therapeutic; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..1443  
 FT /\*tag= a  
 FT /product= TANGO 224  
 XX  
 PN WO200039284-A1.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 23-DEC-1999; 99WO-US31025.  
 XX  
 PR 30-DEC-1998; 98US-0223546.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Holtzman DA;  
 XX  
 XX WPI: 2000-465743/40.  
 DR P-PSDB; AAB01425.  
 DR  
 XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
 PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
 PT arthritis, psoriasis and autoimmune diseases  
 XX  
 PS Claim 1; Fig 7; 209pp; English.  
 XX  
 CC Nucleic acids encoding TANGO polypeptides are useful as modulating  
 CC agents for regulating cellular processes like asthma, graft  
 CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
 CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
 CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
 CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune

CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
 CC lupus erythematosus. The nucleic acids are also useful for producing  
 CC transgenic animals and the TANGO polypeptides themselves. Partial  
 CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
 CC forensic biology, for diagnostic assays, prognostic assays,  
 CC pharmacogenomics and for monitoring clinical trials. TANGO  
 CC polypeptides are suitable for both prophylactic and therapeutic  
 CC methods for treating a subject at risk of a disorder or having a  
 CC disorder associated with aberrant TANGO expression. A wide range  
 CC of cellular disorders can be treated.

XX  
 SQ Sequence 2689 BP; 633 A; 740 C; 734 G; 582 T; 0 other;

# alignment\_scores:

Quality: 4775.00 Length: 874  
 Ratio: 5.520 Gaps: 2  
 Percent Similarity: 98.970 Percent Identity: 98.741

## alignment\_block:

US-10-044-807-2 x AAAA7458 ..

Align seg 1/1 to: AAAA7458 from: 1 to: 2689

```

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17
67 ATGGAATGTCGCCGCGGGCAATCCTCGCACACATGCTCTCTTTCTGGC 116
17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluAlaAspG 34
117 TTTCCGTCCTGAGTTCAGGACCGCACGCTCCGAGGAGCAGCGGAGC 166
34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
167 GCCTGTGGGATGCTCGGGGCCATGGAGTGAATGCTCACGCACCTGCGGG 216
51 GlyClyAlaSerIySerLeuArgArgCysLeuSerSerIySerCysGly 67
217 GGAGGGCCCTCCCTACTCTCTGAGGCGCTGCCTGAGCAGCAAGAGCTGTGA 266
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
267 AGGAAGAAATATCCGATACAGACATGCAGTAATGTGGACTGCCACCAG 316
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
317 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGTCAAG 366
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
367 CACCATGGCCAGTTTATGATGATGGCTTCCCTGTCTTAATGACCCGTGACAA 416
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValGluL 134
417 CCCATGTTTCACTCAAGTGCCAGCCAAAGGAAGAACACCTCGTTGTGAAC 466
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
467 TAGCACCCTAAGGCTTAGATGGTACGCGTTGCTATACAGAAATCTTTGGAT 516
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuGly 167
517 ATGTGCATCAGTGGTTATGCCAAATGTGTGGCTGCATCACCAGCTGGG 566
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
567 AAGCACCCTCAAGGAAGATAACTGTGGGGTCTCAACGGGAGATGGGTCCA 616
184 hrCysArgLeuValArgGlyGlnTyrIySerGlnLeuSerAlaThrLys 200
617 CCTGCGGGCTGGTCCGAGGCGATATAAATCCAGGCTCTCCGCAACCCAAA 666
201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217

```

```

667 TCGGATGATATCTGTGGTTGCAATTCCTATGGAAGTAGACATATTCGCCT 716
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
717 TGTCTTAAAGAGTCTCATCATTATATCTGGAACCAACCAACCTCCAGG 766
234 lyThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250
767 GGACTAAGGTGAACACAGTCTCAGCTCCACAGGAACCTTCTTGTGGAC 816
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
817 AATCTAGTGTGGACTTCCAGAAATTTCCACACAAAGAGATACAGAAAT 866
267 talAGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
867 GGCCTGGACCACTCACAGCAGATTTTCATTTGTCAGATTCGTAACCTCGGGCT 916
284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
917 CCGCTGACAGTACAGTCCAGTTTCATCTTATCAACCATCATCCACCGA 966
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317
967 TGGAGGAGACGGATTTCTTTCTTGTCTGTCAGCAACCTGTGGAGGAGTTA 1016
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
1017 TCAGCTCACATCGCTGAGTGCTACGATCTGAGGAGCAACCGTGTGTG 1066
334 laAspGlnTyrCysHisTyrTrpProGluAsnIleLysProLysProLys 350
1067 CTGACCAATACTGTCACTATTACCAGAGAAACATCAAAACCAACCAAG 1116
351 LeuGlnGluCysAsnLeuAspProCysProAla..... 361
1117 CTTCAGGAGTGCAACTTGGATTCCTGTGTCAGCCAGAGGGTGGCATTAAT 1166
362 .....Ser..AspGlyTyrLysGlnIleMetProTyrAspLeuTyrHi 375
1167 GTTCTTAACAGTGACGGATACAAACAGATCATGCTTATGACCTTACCA 1216
375 sProLeuProArgTrpGluAlaThrProTrpThrAlaCysSerSerSerc 392
1217 TCCCTTCTCCTCGGTGGAGGCGACCCCATGGACGCGTGTCTCTCTCGT 1266
392 ysGlyGlyGlyIleGlnSerArgAlaValSerCysValGluAspIle 408
1267 GTGGGGGGGGCATCCAGCGGGGCGAGTTTCTGTGTGGAGGAGGACATC 1316
409 GlnGlyHisValThrSerValGluGluTrpLysCysMetTyrThrProLy 425
1317 CAGSGGCATGTCACCTTCCAGAGAGTGGAAATGCATGTACACCCCTAA 1366
425 sMetProIleAlaGlnProCysAsnIlePheAspCysProLysTrpLeuA 442
1367 GATGCCCATCGCGAGCCCTGCAACATTTTGTACTGCCCTAAATGGCTGG 1416
442 laGlnGluTrpSerProCysThrValThrCysGlyGlnGlyLeuArgTyr 458
1417 CACAGGAGTGGTCTCCCTGGACAGTGCATGTGGCCAGGGGCTCAGATAC 1466
459 ArgValValLeuCysIleAspHisArgGlyMethHisThrGlyGlyCysSe 475
1467 CGTGTGTCTCTGCATCGACCATCGAGGAATGCACACAGGAGGCTGTAG 1516
475 rProLysThrIySerProHisIleLysGluGluCysIleValProThrProC 492
1517 CCCCCAAACAAAGCCCCACATAAAGAGGAATGCATGTACCCACCTCCCT 1566
492 ysTyrLysProLysGluLysLeuProValGluAlaLysLeuProThrPhe 508

```

```
1567 GCTATTAACCCAAAGAGAAACTTCCAGTCAGGCGCAAGTTGCCATGGTTC 1616
509 LysGlnAlaGlnGluLeuGluGluGlyAlaAlaValSerGluGluProSe 525
1617 AAACAAGCTCAACAGCTAGAGAAGGAGCTGCTGTGTACAGGAGCCCTC 1666
525 rPheIleProGluAlaTrpSerAlaCysThrValThrCysGlyValGlyT 542
1667 GTTCATCCCAAGAGCCCTGTCGGCCCTGCACAGTCACCTGTGGTGGGA 1716
542 hrGlnValArgIleValArgCysGlnValLeuLeuSerPheSerGlnSer 558
1717 CCAGGTGCGAATAGTACGTGCAGGTGCTCTCTCTCTCTCTCTCTCTCT 1766
559 ValAlaAspLeuProIleAspGluCysGluGlyProLysProAlaSerG1 575
1767 GTGGCTGACCTGCTATTGACGAGTGTGAAGGCGCCCAAGCCAGCATCCCA 1816
575 nArgAlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsnP 592
1817 GCGTGGCTGTTATGCAGGCCCATGCAGCGGGGAAATTCCTGAGTTCAACC 1866
592 roAspGluThrAspGlyLeuPheGlyGlyLeuGlnAspPheAspGluLeu 608
1867 CAGCGAGACAGATGGCTCTTTGGTGGCTCGAGGATTCGAGAGCTG 1916
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG1 625
1917 TATGACTGGGAGTATGAGGGGTTCCACCAAGTCTCCGAGTCCGTGCGAGG 1966
625 yGlyValGlnGluAlaValValSerCysLeuAsnLysGlnThrArgGluP 642
1967 AGGTGTCCAGGAGGCTGTGCTGAGCTGCTTGAACAAACACAGACTCGGAGC 2016
642 roAlaGluGluAsnLeuCysValThrSerArgArgProGlnLeuLeu 658
2017 CTGCTGAGGAGAACCTGTGGCTGACACAGCGCGCGCCCCACAGCTCCTG 2066
659 LysSerCysAsnLeuAspProCysProAlaArgTrpGluIleGlyIysTr 675
2067 AAGTCTGCAATTTGATCCCTCCCAAGCAAGGTGGGAAATTCGCAAGTG 2116
675 pSerProCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValP 692
2117 GAGTCCATGTAGTCTCACATGTGGGTGCGGCTTACAGACCAGAGCGTCT 2166
692 heCysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAla 708
2167 TCTGCAGCCACCTGCTTCCAGAGATGATGATGAACAGTCATCCCGGCT 2216
709 AspGluLeuCysArgGlnProLysProSerThrValIleAlaCysAsnAr 725
2217 GATGAGCTGTGTGCGCCAGCCCAAGCCAGCACGCTGCAAGCTTGTAAACC 2266
725 qPheAsnCysProProAlaTrpTyrProAlaGlnTrpGlnProCysSerA 742
2267 CTTTAATGCCCCAGCCCTGTACCTGTCACAGTGCAGCGCTGTTCCTCA 2316
742 rgThrCysGlyGlyGlyValGlnLysArgGluValLeuCysLysGlnArg 758
2317 GAACGTGTGGCGGGGTGTTCAGAAACGTGAGGTCTTTGCAAGCAGCGC 2366
759 MetAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlaSe 775
2367 ATGGCTGATGGCAGCTTCCTGGAGCTTCCTGACACCTTCGTGTTTCAGTTC 2416
775 rLysProAlaCysGlnGluAlaCysLysLysAspAspCysProSerGluT 792
2417 AAAACCTGGCTGCCAAGCAGCATGCAAGGAAGATGACTGTCCCGAGGAGT 2466
792 rPheLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyThr 808
2467 GGGTCTCTCAGACTGGACAGAGTGTTCACAAGCTGCGGGGAAGGCACC 2516
```

```
809 GlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerTh 825
2517 CAGACTCGAAGCGCAATTCGCCGAAAGATGCTGAAACCGGCATCTCAAC 2566
825 rValValAsnSerThrLeuCysProProLeuProPheSerSerSerIleA 842
2567 GGTGTCAATTCACCCCTGTGCCCGCCCTGCTTCTCTCTCTCTCTCTCT 2616
842 rgProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLys 858
2617 GGCCTGTATGTGGCAACCTGTGCAAGGCCCGGGCGGCATCCACGAAG 2666
859 HisSerProHisIleAlaAla 865
2667 CACAGCCCGCACATCGCGGCC 2687

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAD14364
seq_documentation_block:
ID AAD14364 standard; cdna; 5076 BP.
XX
AC AAD14364;
XX
DT 01-NOV-2001 (first entry)
XX
DE Novel human protein (NHP) cdna #1.
XX
KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;
cytostatic; antisense therapy; angiogenesis; biological disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..5076
FT FT /*tag= a
FT FT /product= *Novel human protein*
XX
PN WO200161011-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05290.
XX
PR 17-FEB-2000; 2000US-0183282.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
Sands AT;
XX
DR WPI: 2001-514776/56.
DR P-PSDB; AAE07863.
XX
PT An isolated nucleic acid encoding a thrombospondin useful as a
hybridization probe and gene therapy treatments of cancer -
XX
PS Claim 1; Page 29-30; 56pp; English.
XX
CC The invention relates to novel human polynucleotides encoding proteins
that share sequence similarity with animal proteins having thrombospondin
repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,
prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and
gene trapped human cells. NHPs are thrombospondins useful for treating
biological disorders involving angiogenesis, cancer and development and
also in pharmacogenomic applications. NHPs are useful as a hybridisation
probe for screening libraries, assessing gene expression patterns and
also in gene therapy. Proteins having thrombospondin repeats act as
receptors, secreted extracellular matrix proteins and proteases. The
present sequence is a cDNA encoding novel human protein (NHP).
XX
SQ Sequence 5076 BP; 1384 A; 1227 C; 1338 G; 1127 T; 0 other;
```

alignment\_scores:  
 Quality: 3956.50 Length: 1775  
 Ratio: 3.248 Gaps: 31  
 Percent Similarity: 68.620 Percent Identity: 43.324

alignment\_block:  
 US-10-044-807-2 x AAD14364 ..

Align seg 1/1 to: AAD14364 from: 1 to: 5076

```

22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 ACCTCAAGAAACACTCGTTCAGATGAAGCAAAAGATGGCAACTGGGATGC 239
38 aTrpGlyProTrpSerGluCysSerArgThrCysGlyGlyGlyAlaSerT 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
240 TTGGGGCGACTGGAGTACGTCTCCCGACCTGTGGGGGAGGAGCATCAT 289
55 yrSerLeuArgCysLeuSerSerLysSerCysGluGlyArgAsnIle 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290 ATTCTCTGGGAGATGTTGACTTGAAGGAATGTGAAGGGCAGAAATTT 339
72 ArgTyrArgThrCysSerAsnValAspCysProProGluAlaGlyAspPh 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
340 CGGTCAAGACATGCGACCAATCATGACTGCCCTCCAGATGCAGAAAGATTT 389
88 eArgAlaGlnGlnCysSerAlaHisAsnAspValLysHisHisGlyGlnP 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
390 CAGAGCCACGACGTCTCAGCTACAAATGATGTCAGATACAGGGGCATTT 439
105 heTyrGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
440 ACTATGAATGGCTTCCACGATATAATGATCTCTGCTGCCCGTGTGCACTC 489
122 LysCysGluAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
490 AAGTGTATGCACGACGACAAACTTGGTGGTGGAGTGGCACTTAAGGT 539
138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
540 ACTGGATGGAACTCGTGCAACACGGACCTCTTGGACATGTATCAGTG 589
155 lyLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLys 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
590 GCATCTGTGAGGACAGTGGGTGGATCGGCACTGGGAAGCAATGCCAAG 639
172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
640 GAGGACAACTGTGGATCTGTGTCGGCGGATGGCTCCACCTGCGAGGCTGT 689
188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspThrY 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
690 ACGGGACAATCAAGTCAACAGTCTCTCTCTGAAAAAAGAGAGAAAAATG 739
205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
740 TAATGTCTGTCTCTTGGGAAGTCTGAAGTGTGAGAAATTACAGTGAAGGA 789
222 proAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyG 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
790 CTTGCCCACTCTTTATTTGATCAAAACAACTTCAAGGAAGCAAGGAGA 839
238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerValA 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
840 ACACAGCTTTAACAGCCCGCGCTCTTGTCTGTAGAAAAACACACAGTGG 889
255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
890 AATTTCAGAGGGGCTCCGAGGACAACTTTTAAGATTCAGGACCTCTG 939
272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

940 ATGGCTGATTTTCATCTTCAAGACCAGGTACACTGCAGCCAAAGACAGCGT 989
288 rValGlnPheIlePheTyrGlnProIleIleHisArgTyrArgGluThrA 305
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
990 GGTTCAGTTCTTCTTTACCAGCCCATCAGTCATCAGTGGAGACAAACTG 1039
305 spPhePheProCysSerAlaThrCysGlyGlyGlyTyrGlnLeuThrSer 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1040 ACTTCTTTCCTGCTGACGTGTGAGGTGTGAGGAGGTATCAGTCAATCT 1089
322 AlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1090 GCTGAATGTGGATATCCCGCTTGAAGAGGTAGTCTCTGACCAATTATG 1139
338 sHisTyrTyrProGluAsnIleLysProLysProLysLeuGlnGluCysA 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1140 TCACTACTACCCIGAAATGTAAAAACCAAAACCAAACTGAAGGAATGCA 1189
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1190 GCATGGATCCCTGCCCATCAAGTGTGATGATTTAAAGAGATAATGCCCTAT 1239
372 AspLeuTyrHisProLeuProArgTyrGluAlaThrProTyrThrAlaCy 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1240 GACCACTTCAACCTCTTCCTCGCTGGAAACATTAATCTTGGACTGCATG 1289
388 sSerSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerCysValG 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1290 TTCGTGCTCTGTGGAGGAGGATTCAGAGACGAGGCTTTTGTGTGTAG 1339
405 luGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysMet 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1340 AGAATCCATGCATGAGAGATATTCAGGTGGGAAGTAATGGAAGTGCATG 1389
422 TyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCysPr 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1390 TACCACCCAAACCCCAAGTTATGCAACTTGTAACTCTTTTGTGATGCC 1439
438 oLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysGlyGlnG 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1440 CAAGTGGATGGCATGGAGTGTCTCATGTCACAGTGACITGTGGCCGAG 1489
455 lyLeuArgTyrArgValValLeuCysIleAspHisArgGlyMetHisThr 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1490 GGTATACGGTACCGGGTGTCTTGTGTATTAACCAACCGCGAGAGCATGT 1539
472 GlyGlyCysSerProLysThrLysProHisIleLysGluGluCysIleVa 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1540 GGGGCTGCATCCCACTGAAGTTACACATCAAGAAGAAATGTGTCTAT 1589
488 lProThrProCysTyrLysProLysGluLysLeuProValGluAlaLysL 505
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1590 TCCCATCCCGTGTATAAACCAAAAGAAAAAGTCCAGTGGAGCAAAAT 1639
505 euProTrpPheLysGlnAlaGlnGluLeuGluGlyAlaAlaValSer 521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1640 TGCCTTGGCTGAAACAAGACACAAAGACTAGAAAGACCAAGAAATAGCAACA 1689
522 GluGluProSerPheIleProGluAlaTrpSerAlaCysThrValThrCy 538
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1690 GAAGAACCAAGTTTCATTCAGAACCTTGTTCAGCTGCAGTACACAGTG 1739
538 sGlyValGlyThrGlnValArgIleValArgCysGlnValLeuLeuSerP 555
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1740 TGGGCCAGGTGTGAGGTCCCGAGGTGAAGTGCCTGTGTCTCTCATAT 1789
555 heSerGlnSerValAlaAspLeuProIleAspGluCysGluGlyProLys 571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1790 TCAGCAGACTGAGACTGAGTCCCGGAGGAGAGTGTGAAGGCCCAAG 1839
572 ProAlaSerGlnArg.....AlaCysTyrAlaGlyProCy 583
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1840 CTGCCACCGAAGCGCCTCTCTCTGGAACGATGTGATGAGAGCCCGGC 1889

```



583 sSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheG 600  
1890 CTCCTCCGAGAGCTACACATCCCTCCCT..... 1917  
600 LyGlyLeuGlnAspPheAspGluLeuThrAspTrpGluThrGluGlyPhe 616  
1918 .....GAGGACAGTACAGACACATACAGCTGGAGTACGCTGGGTTT 1959  
617 ThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValValse 633  
1960 ACCCTTGCACAGCAACATCTTGGGAGGCCATCAAGAACCCATAGCAGT 2009  
633 rCysLeuAsnLysGlnThrArgGluProAlaGluAlaAsnLeuCysValt 650  
2010 GTGCTTACATATCAGACCCAGCAGACAGTCAATGACAGCTGTGTGATA 2059  
650 hrSerArgArgProGlnLeuLeuLysSerCysAsnLeuAspProCys 666  
2060 TGTCTCCAGCGTCTCCAGCCATGAGCAGGCCCTGTACACAGAGCCCTGT 2109  
667 ProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCysG1 683  
2110 CCCCCAGGTGGCATGTGGCTCTTGGGGCCCTGCTCAGCTACCTGTGG 2159  
683 yValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerArgG 700  
2160 AGTTGAATTCAGACCCGAGATGTACTGCTGTCAC..... 2196  
700 LuMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnProLys 716  
2197 ..CCAGGGAGACCCCTGCCCTCTGGAGAG...TGCCGAGATGAAGAAG 2241  
717 ProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTrpTy 733  
2242 CCCCATGCTTTACAGCATGCAATCATGTTGACTGCGCTCCTGGCTGGCA 2291  
733 rProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnL 750  
2292 CATTGAAGAAATGGCAGCAGTGTTCAGGACTTGTGGGGGGAACTCAGA 2341  
750 ysArgGluValLeuCysLysArgMetAlaAspGlySerPheLeuGlu 766  
2342 ACAGAGAGTCACTCTGGCAGCTGCTACAGGATGCGACCTTTTGAAT 2391  
767 LeuProGluThrPheCysSerAlaSerLysProAlaCysGlnAlaCy 783  
2392 CTCTCAGATGAATGTGCCAAGGACCCAGGCATGCTCTCACAAAGTCTG 2441  
783 sLysLysAspCysProSerClnTrpLeuLeuSerAspTrpThrGluC 800  
2442 TGCCAGGACAGACTGCTCTCCACATTTAGCTGTGGGAGACTGGTCAAGT 2491  
800 ysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCysArg 816  
2492 GTTCTGTCAAGTGTGGTGTGAATCCAGAGAGAAGACAGGTGTGTCAA 2541  
817 LysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCysPr 833  
2542 AGGCTGGCAGCCAAAGGTCGGGCGCATCCCTCAGTGAGATGTGTCAG 2591  
833 oProLeuProPheSerSerIleArgProCysMetLeuAlaThrCysA 850  
2592 GGATCTACCAAGGGTCCCTCTGTGAAGATCTTGCCAGATGCCTGAGTGCA 2641  
850 laArg...ProGlyArgProSerThrLys.....HisSerProHIs 862  
2642 GTAATCAATCAATCAGATGACAGACAAAATTTGGTCAGCAGGTCGCCAG 2691  
863 IleAlaAlaArgLysValTyrIleGlnThrArgGlnArgLysLe 879  
2692 ATCCTCAGTGTCCAGAGATCTACATTCAGACAAAGGAAGAGAGCGTAT 2741

879 uHisPheValValGlyGlyPheAlaTyrLeuLeuProLysThrAlaValV 896  
2742 TAACCTGACCATTGGTAGCAGAGCTATTGCTGCCAACACATCCGTGA 2791  
896 alLeuArgCysProAlaArgValArgLysProLeuIleThrTrpGlu 912  
2792 TTATTAAGTGGCCGCTGGAGATTCCAGAAATCTCTGATCCAGCTGGGAG 2841  
913 LysAspGlyGlnHisLeuSerSerThrHisValThrValAlaProPh 929  
2842 AAGATGGCGTTCCTGCGAGAACTCCAAACGGCTTGCGATCACCAGTC 2891  
929 eGlyTyrLeuLysIleHisArgLeuLysProSerAspAlaGlyValTyr 946  
2892 AGGCTCACTAAATCCACGGTCTTGCTGCCCGCACATCGCGCTGTACC 2941  
946 hrCysSerAlaGlyProAlaArgGluHisPheValIleLysLeuIleGly 962  
2942 GGTGCTATGCGGCTCTGCGAGAAACAGTTGTCTCAAGCTCATTTGT 2991  
963 GlyAsnArgLysLeuValAlaArgProLeuSerProArgSerGluGluG 979  
2992 ACTGACAAACCGCTCATCGCACGCCCA..... 3018  
979 uValLeuAlaGlyArgLysGlyProLysGluAlaLeuGlnThrHisL 996  
3019 .....GCCCTCAGGGAGCCTATGAGGGAATATCTGGGATGGACACACA 3061  
996 yHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGluLysArgGly 1012  
3062 GCGAAGCAATAGTTTGGGAGTACATGGCACAAA.....ATGAGGCAA 3105  
1013 LeuAlaAlaAsnProGlySerArgTyr.....AspAspLeuValSerAr 1027  
3106 ATGTGGNAATAACAAAATGACCTTATCTCGATGATGACCATATTAGTAA 3155  
1027 g.....LeuLeuGluGlnGlyTyrProGlyGluL 1038  
3156 CCAGCCTTCTTGAGAGCTCTGTAGGCCACTGCAGCAATCTCCAGGAA 3205  
1038 euLeuAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrSer 1054  
3206 GCACCAACTCTCTGGAGTTGAAGAAATAGCAGTTTGAAGCAGCAGTTAA 3255  
1055 GluGluAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMe 1071  
3256 CAA.....GGAGCA.....TATAGCAT 3272  
1071 tValThrGluGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGlnG 1088  
3273 GGATACAGCCAG.....TTTGATGAGCTGATGAAGAAACATGAGTCAGC 3316  
1088 InProGlu.....GluLeuArgAspLeuTyrSerLysHisLeuValAla 1102  
3317 TCATGAAACCCGAGAGGTGACGCGATGATCTTGCTCCCGAGTGTATAT 3366  
1103 GlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspThrIe 1119  
3367 CAGCTGGTGGCCCAATTACCAAGGCACAGCCCAACACATGACAG..... 3411  
1119 uLeuLysProSerGluArgThrSerProValThrLeuSerProHisL 1136  
3412 .TGGCGGGGCATCCAGGAAGAGACCTCCT..... 3441  
1136 yHisValSerGlyPheSerSerSerLeuArgThrSerSerThrGlyAsp 1152  
3442 .....GCTGCTCAGCTCAGA.....GGGGAA 3462  
1153 AlaGlyGlyCysArgArgProHisArgLys..... 1163  
3463 ACAGGGAGTGTGCCCAAGCTCGCATGCAAAAACACTCAGGCAAGCTGAC 3512  
1164 .....ProThrIleLeuArgLysIleSerAlaAlaGlnG 1175

```

3513 ATTCAAGCCCAAGGACCTGTTCTCATGAGCAA...AGCCAAACCTCCCT 3559
||||:||||:||||:||||:||||:
1175 lnLeuSerAlaSerGluValVal...ThrHisLeuGlyGlnThrValAla 1190
||||:||||:||||:||||:||||:||||:||||:||||:
3560 CAATTTCAATTAATAAATAAATAAATCCAGGATTCGAAATACAGATATAC 3609
||||:||||:||||:||||:||||:||||:||||:||||:
1191 LeuAlaSerGlyThrLeuSerValLeuLeuLeuHisCysGluAlaIleGlyHI 1207
||||:||||:||||:||||:||||:||||:||||:||||:
3610 ATTACAAAAGGACAGAGGTCATCATATATACTGTGTGACCTTATTACCCC 3659
||||:||||:||||:||||:||||:||||:||||:||||:
1207 sProArgProThrIleSerTrpAlaArgasnGlyGluGluValGlnPheS 1224
||||:||||:||||:||||:||||:||||:||||:||||:
3660 CAGTGAGGCCACATATACATGAGCCAGGATGGAACCTTGTTCACAGCCCT 3709
||||:||||:||||:||||:||||:||||:||||:||||:
1224 erAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaPro 1240
||||:||||:||||:||||:||||:||||:||||:||||:
3710 CAGTAAATAATTTGGATGGAACTCGGGAAGATACAGATACAGAAATCCT 3759
||||:||||:||||:||||:||||:||||:||||:||||:
1241 ValGluAlaAspValGlyPheTrpThrCysAsnAlaThrAsnAlaLeuGI 1257
||||:||||:||||:||||:||||:||||:||||:||||:
3760 ACAAGAAAGAACAGGACATATATGAATGTTCTGTAGCTAATCATCTTGG 3809
||||:||||:||||:||||:||||:||||:||||:||||:
1257 yTrpAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
||||:||||:||||:||||:||||:||||:||||:||||:
3810 TTCAGATGTGGAAGTCTCTCTGTCTGTATGACAGAGCCACCTGTCTATCT 3859
||||:||||:||||:||||:||||:||||:||||:||||:
1274 ysthrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
||||:||||:||||:||||:||||:||||:||||:||||:
3860 TGCTCTTGAAAGAAATATACCAAAACAGACGACACCAATCTGTCTGTT 3909
||||:||||:||||:||||:||||:||||:||||:||||:
1291 AspIleGlySerThrIleLysThrValGlnGlyValAsnValThrIleAs 1307
||||:||||:||||:||||:||||:||||:||||:||||:
3910 GTGGTTGGAGGATCTGGAGGAGCCCTTGGACCAACGTGCAATCCG 3959
||||:||||:||||:||||:||||:||||:||||:||||:
1307 nCysGlnValAlaGlyValProGluAlaGluValThrTrpPheArgAsnL 1324
||||:||||:||||:||||:||||:||||:||||:||||:
3960 ATGTCCTGTAAAAGGTGCTCCCTCAGCCATAATATACCTTGGTTGAAGAG 4009
||||:||||:||||:||||:||||:||||:||||:||||:
1324 ysSerLysLeu...GlySerProHisHisLeuHisGluGlySerLeuLeu 1339
||||:||||:||||:||||:||||:||||:||||:||||:
4010 GAGGATCTCTGAGTGGCAATGTTCCCTTGCTTTTCAATGATCCCTGTTG 4059
||||:||||:||||:||||:||||:||||:||||:||||:
1340 LeuThrAsnValSerSerSerAspGlnGlyLeuTrpSerCysArgAlaAl 1356
||||:||||:||||:||||:||||:||||:||||:||||:
4060 TTGCAGAAATTTTCCCTTGAATATGAGGAACCTACGCTCTGCATAGCCAC 4109
||||:||||:||||:||||:||||:||||:||||:||||:
1356 aAsnLeuHisGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspP 1373
||||:||||:||||:||||:||||:||||:||||:||||:
4110 CAATGCTTGGAAAGGACGTGGCAACATCTGTACTCCACTTGTCTGGAAC 4159
||||:||||:||||:||||:||||:||||:||||:||||:
1373 roProGlnValProThrGlnLeuGluAspIleArgAla... 1385
||||:||||:||||:||||:||||:||||:||||:||||:
4160 GAAGATGGCCAGAGAGTAGAATCGTATTTCTGCAAGGACATAAAAAGTAC 4209
||||:||||:||||:||||:||||:||||:||||:||||:
1386 LeuLeuAlaIaIaThrGlyProAsnLeuProSerValLeuThrSerProLe 1402
||||:||||:||||:||||:||||:||||:||||:||||:
4210 ATTCTCAGGCAACCAACACAGTAAACCAACAGC... 4242
||||:||||:||||:||||:||||:||||:||||:||||:
1402 uGlyThrGlnLeuValLeuLeuAspProGlyAsnSerAlaLeuLeuGlyCysP 1419
||||:||||:||||:||||:||||:||||:||||:||||:
4243 .....AATGACCCA..... 4251
||||:||||:||||:||||:||||:||||:||||:||||:
1419 roIleLysGlyHisProValProAsnIleThrTrpPheHisGlyGlyGln 1435
||||:||||:||||:||||:||||:||||:||||:||||:
4252 .....ACAGGAGAACCCCGCCT..... 4269
||||:||||:||||:||||:||||:||||:||||:||||:
1436 ProlleValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGlyGI 1452
||||:||||:||||:||||:||||:||||:||||:||||:
4269 ..... 4269
||||:||||:||||:||||:||||:||||:||||:||||:
1452 nIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyLupheSerC 1469
||||:||||:||||:||||:||||:||||:||||:||||:

```

```

4269 ..... 4269
||||:||||:||||:||||:||||:||||:||||:||||:
1469 ysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuVal 1485
||||:||||:||||:||||:||||:||||:||||:||||:
4269 ..... 4269
||||:||||:||||:||||:||||:||||:||||:||||:
1486 IleGlnAspTrpTrpTrpSerValAspArgLeuAlaThrCysSerAlaSe 1502
||||:||||:||||:||||:||||:||||:||||:||||:
4270 ...CAAGAGCCTTTTGGAGCCTGTAACATGGTCACATTTGTTCTGCCAC 4316
||||:||||:||||:||||:||||:||||:||||:||||:
1502 rCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLeu...A 1518
||||:||||:||||:||||:||||:||||:||||:||||:
4317 CTGTGGTCATTTGGGAGCCCGCATTCAGAGACCCAGTCGTGTGTATGCCA 4366
||||:||||:||||:||||:||||:||||:||||:||||:
1518 snSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProAla 1534
||||:||||:||||:||||:||||:||||:||||:||||:
4367 ATGGCAGGAAGTGAAGTGAAGCCCTGTGTATCACCTCCAGAACCCACTG 4416
||||:||||:||||:||||:||||:||||:||||:||||:
1535 ValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMetVa 1551
||||:||||:||||:||||:||||:||||:||||:||||:
4417 GCTGGGTTTGAAGCCTGTAAATCCGGGACTGCCAGCGAGGTGGTTCAC 4466
||||:||||:||||:||||:||||:||||:||||:||||:
1551 lThrSerTrpSerAlaCysThrArgSerCysGlyGlyValGlnThrA 1568
||||:||||:||||:||||:||||:||||:||||:||||:
4467 AAGTGTGTGTTCACAGTCTGTCTGTCTGCGGTGAAGGATACCACAGTC 4516
||||:||||:||||:||||:||||:||||:||||:||||:
1568 rgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProVal 1584
||||:||||:||||:||||:||||:||||:||||:||||:
4517 GGCAGGTGACGTGCAAGCGGCAAAAGCAATGGAACCTGTGCAGGTGGTG 4566
||||:||||:||||:||||:||||:||||:||||:||||:
1585 SerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGlnAl 1601
||||:||||:||||:||||:||||:||||:||||:||||:
4567 TCTCAAGAGCATGTGCCCT...AAAGACCGGCTCTGGGAAGAAAC 4613
||||:||||:||||:||||:||||:||||:||||:||||:
1601 acYsAsnGlnGlnLeuCysValGluTrpAlaPheSerSerTrpGlyGlnC 1618
||||:||||:||||:||||:||||:||||:||||:||||:
4614 ATGTTTTGGTCATCCATGTTCAAGTGG.....GAACCGGGAACCGGT 4657
||||:||||:||||:||||:||||:||||:||||:||||:
1618 ysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnVal 1634
||||:||||:||||:||||:||||:||||:||||:||||:
4658 GTCTTGGACGTTGCATGGCGGTGCTGTGAGGATGTCAGCAGCTCACACA 4707
||||:||||:||||:||||:||||:||||:||||:||||:
1635 PheCysGlnThrArgAspGlyIleThrLeuProSerGluGlnCysSerAl 1651
||||:||||:||||:||||:||||:||||:||||:||||:
4708 GCTTGTCAACACACAGCTCT.....GACTCCAACTGTGTATGA 4745
||||:||||:||||:||||:||||:||||:||||:||||:
1651 aLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysSerV 1668
||||:||||:||||:||||:||||:||||:||||:||||:
4746 CAGAAAGAGACCCACCITTAAGAGGAACTGCACATCAGGGCCCTGTGATG 4795
||||:||||:||||:||||:||||:||||:||||:||||:
1668 alHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyAsn 1684
||||:||||:||||:||||:||||:||||:||||:||||:
4796 TGTGTGGCACACAGGCCCTTGGAAAGCCTGTACAGCAGCCTGTGGCAG 4845
||||:||||:||||:||||:||||:||||:||||:||||:
1685 TyrGlyPheGlnSerArgValGluCysValHisAlaArgThrAsnLy 1701
||||:||||:||||:||||:||||:||||:||||:||||:
4846 ...GGTTTCCAGTCTCGAAAGTCGACGTATCCACACAAAGGAGTTGCA 4892
||||:||||:||||:||||:||||:||||:||||:||||:
1701 sAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpG 1718
||||:||||:||||:||||:||||:||||:||||:||||:
4893 ACCTGTGCCCAAGAGACACTGTGTACAGAAAAGAAACCAATTTCTGGC 4942
||||:||||:||||:||||:||||:||||:||||:||||:
1718 lnArgCysAsnIleThrProCysGluAsnMetGluCysArgAspThrThr 1734
||||:||||:||||:||||:||||:||||:||||:||||:
4943 GGCACGTGT...CTTGGGCCCTCCTGTGTATAGAGACTGCACAGACACA 4989
||||:||||:||||:||||:||||:||||:||||:||||:
1735 ArgThrCysGluLysValLysGlnLeuLysLeuCysGlnLeuSerGlnPh 1751
||||:||||:||||:||||:||||:||||:||||:||||:
4990 CACTACTGTATGTTGTAAACATCTTAATTTGTTGTTCTTAGACCCGCTA 5039
||||:||||:||||:||||:||||:||||:||||:||||:
1751 eLysSerArgCysCysGlyThrCys 1759
||||:||||:||||:||||:||||:||||:||||:||||:
5040 CAAACAAAGGTGCTGCCAGTCATGT 5064
||||:||||:||||:||||:||||:||||:||||:||||:

```

seq\_name: /STDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT: AAD17757

seq\_documentation\_block:

ID AAD17757 standard; DNA; 7260 BP.

XX AC AAD17757;

XX DT 10-DEC-2001 (first entry)

XX DE Human novel KIAA1233-like protein, NOV-2a encoding DNA.

XX KW Human: NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;  
 KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;  
 KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;  
 KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;  
 KW cytostatic; nootropic; anti-infertility; cancer; chromosome 15;  
 XX NOV-2a protein; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT 5'UTR 1..135

FT /\*tag= a

FT CDS 136..5211

FT /\*tag= b

FT /product= "Human novel KIAA1233-like protein, NOV-2a"

FT sig\_peptide 136..213

FT /\*tag= c

FT mat\_peptide 214..5208

FT /\*tag= d

FT /product= "Human mature novel KIAA1233-like protein,  
 NOV-2a"

FT 3'UTR 5212..7260

FT /\*tag= e

XX WO200162928-A2.

XX PD 30-AUG-2001.

XX PF 26-FEB-2001; 2001WO-US06151.

XX PR 25-FEB-2000; 2000US-0184951.

XX PR 28-FEB-2000; 2000US-0185548.

XX PR 01-MAR-2000; 2000US-0185967.

XX PR 18-APR-2000; 2000US-0197723.

XX PR 27-APR-2000; 2000US-0199957.

XX PR 23-FEB-2001; 2001US-0789390.

XX PA (CURA-) CURAGEN CORP.

XX PI Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;

XX WPI; 2001-582051/65.

XX P-PSDB; AAE10609.

XX New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like

XX polypeptide for diagnosing and treating pathological disorders, such as

XX Parkinson's disease and for use in pharmacogenomics -

XX Claim 9; Page 20-22; 189pp; English.

XX The invention relates to novel human polypeptides referred as NOV-X

XX and their corresponding nucleic acid sequences. NOV-X collectively

XX include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like

XX polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel

XX STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e

XX which are novel trypsin inhibitor-like polypeptides. NOV-X is used to

XX identify a potential therapeutic agent that can modulate its activity

XX and can be used for treating a pathology related to aberrant expression

XX or aberrant physiological interactions of NOV-X. NOV-X or its DNA is

XX used to determine the presence or predisposition to a disease associated

CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used  
 CC to treat or prevent a pathology associated with NOV-X. The pathological  
 CC states that can be treated or prevented are haematopoietic, cancer,  
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and  
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and  
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for  
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present  
 CC sequence is a DNA encoding human novel KIAA1233-like protein, NOV-2a.  
 CC NOV-2a gene is located at chromosome 15.

XX SQ Sequence 7260 BP; 2022 A; 1593 C; 1811 G; 1834 T; 0 other;

alignment\_scores:

Quality: 3953.50 Length: 1775

Ratio: 3.246 Gaps: 31

Percent Similarity: 68.620 Percent Identity: 43.268

alignment\_block:

US-10-044-807-2 x AAD17757 ..

Align seg 1/1 to: AAD17757 from: 1 to: 7260

22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38

325 ACCTCAGCAAAACACTCGTTTCAGATGACACAAAGATGCACTGGGATGC 374

38 aTrGlyProThrSerGluCysSerArgThrCysGlyGlyAlaSerT 55

375 TTGGGGCGACTGGAGTGTCTCCCGACCTGTGGGGGAGGAGCATCAT 424

55 YTSerLeuArgArgCysLeuSerSerCysGluGlyArgAsnIle 71

425 ATTCCTCGCGAGATGTTGACTGGAAGGAATTGGAAGGCGCAACATT 474

72 ArgTyrArgThrCysSerAsnValAspCysProGluAlaGlyAspPh 88

475 CGGTACAAGACATGCAGCAATCATGCTCCCTCCAGATGCAGAGATT 524

88 eArgAlaGlnCysSerAlaHisAsnAspValLysHisGlyGlnP 105

525 CAGAGCCCGCAGTGCCTCAGCTCAATGATGCTCAGTATCAGGGCAT 574

105 heTyrGluTrpLeuProValSerAsnAspProAspAspProCysSerLeu 121

575 ACTATGAATGGCTTCCAGTATATGATCTGCTGCTGCTGCTGCTGCT 624

122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138

625 AAGTGTCTATGCACAAAGCAAACTTGTGTGTGGAGCTGGCACCTAAG 674

138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155

675 ACTGGATGGAACTCGTTGCAACACGGACTCCTTGGACATGTTATCAG 724

155 lLeuCysGlnIleValCysAspHisGlnLeuGlySerThrValLys 171

725 GCATCTGTGAGGAGTGGCTGCGATCGCACTGGCACTGGCACTGGCA 774

172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188

775 GAGCACAACTGTGAGTCTGTGCGGGGATGGCTCCACCTGCAGGCTGT 824

188 lArgGlyGlyThrLysSerGlnLeuSerAlaThrLysSerAspThrV 205

825 ACGGGGCAATCAAGTCAACGCTTCTCTCGAAAAAGAGAGAAAAATG 874

205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221

875 TAATTGCTGTTCTTTGGGAGTGGAACTGTGAGAAATACAGTGAAGA 924

222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGly 238



```

833 oProLeuProPheSerSerIleArgProCysMetLeuAlaThrCysA 850
2727 GGATCTACAGGGTCCCTCTGTTGAAGATCTTGCCAGATCCCTCAGTGCA 2776
850 laArg...ProGlyArgProSerThrIys.....HisSerProHis 862
2777 GTAAATCAATCAGATGAGACAAACTTGGTGAGCAGGGTCCGAG 2826
863 IleAlaAlaAlaArgLysValTyrIleGlnThrArgArgGlnArgLysLe 879
2827 ATCCTCAGTGTCCAGAGTCTACATTCAGACAAGGGAAGAGACGTAT 2876
879 uHisPheValValGlyGlyPheAlaTyrLeuLeuProLysThrAlaVal 896
2877 TAACCTGACCATGGTAGCAGAGCTATTGCTGCCAACACATCCCTGA 2926
896 alLeuArgCysProAlaArgArgValArgLysProLeuIleThrTrpGlu 912
2927 TTATTAAGTGCCTGCGAGATTCAGAAATCTCTGATCCAGTGGAG 2976
913 LysAspGlyGlnHisLeuIleSerSerThrHisValThrValAlaProPh 929
2977 AAGGATGCCGTGGCTGCGACAGACTCCAAACCGGCTTGGCATCCCAAGTC 3026
929 eGlyTyrLeuLysIleHisArgLeuLysProSerAspAlaGlyValTyr 946
3027 AGGCTCACTAAATCCACGGCTTGCTGCCCGCAGCATCGGGCTGACC 3076
946 hrCysSerAlaGlyProAlaArgGluHisPheValIleLysLeuIleGly 962
3077 GGTGCATTCAGAGCTCTGCGACAGAAACAGTTGTGCTCAGATCATTTGGT 3126
963 GlyAsnArgLysLeuValAlaAlaArgProLeuSerProArgSerGluGlu 979
3127 ACTGCAACCGGCTCAFCGACGCCCA..... 3153
979 uValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeuGlnThrHisL 996
3154 .....GCCCTCAGGAGGCTATGAGGGAATATCCTGGGATGGACACA 3196
996 yHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGluLysArgGly 1012
3197 GCGAAGCCATAGTTGGGAGTCACATGGGCACAA.....ATGAGGCAA 3240
1013 LeuAlaAlaAsnProGlySerArgTyr.....AspAspLeuValSerAr 1027
3241 ATGTGGAATAACAAATATGACCTTTATCTGGATGATGACCACATTAGTAA 3290
1027 g.....LeuLeuGluGlnGlyGlyTrpProGlyGluL 1038
3291 CCAGCCTTCTTGAGAGCTCTGTAGGCCACTGCACCAATCTGCAGGAA 3340
1038 euLeuAlaSerTrpGluAlaGlnAspSerAlaGluArgAsnThrThrSer 1054
3341 GCACCAACTCCTGGGAGTTGAAGAATAAGCAGTTTGAAGCAGCAGTAA 3390
1055 GluGluAspProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMe 1071
3391 CAA.....GGAGCA.....TATAGCAT 3407
1071 tValThrGluGlnArgArgLeuAspIleLeuGlyAsnLeuSerGlnG 1088
3408 GGATACAGCCAG.....TTTGATGAGCTGATGAAGAAACATGATGATGAT 3451
1088 lnProGlu.....GluLeuArgAspLeuTyrSerLysHisLeuValAla 1102
3452 TCATGGAACCGGAGAGGTGAGGATGATCTTCGCTCCAGCTGATATAT 3501
1103 GlnLeuGlnGlnThrPheArgSerHisLeuGluHisGlnAspThrLe 1119
3502 CAGCTGGTGGCCGAATTAGCCAAAGGCACGCCAACACATGTCAG..... 3546
1119 uLeuLysProSerGluArgArgThrSerProValThrLeuSerProHisL 1136

```

```

3547 .TGGCGGGGCATCCAGGAAGAGACACCTCT..... 3576
1136 yHisValSerGlyPheSerSerLeuArgThrSerSerThrGlyAsp 1152
3577 .....GCTGCTCAGCTCAGA.....GGGAA 3597
1153 AlaGlyGlySerArgProHisArgLys..... 1163
3598 ACAGGAGTGTCTCCCAAGCTCGATGCAAAAACTCAGGCAAGCTGAC 3647
1164 .....ProThrIleLeuArgLysIleSerAlaAlaGlnG 1175
3648 ATTCAAAGCCGAAAGACCTGTCTCATGAGCAA...AGCCAACCTCCCT 3694
1175 lnLeuSerAlaSerGluValVal...ThrHisLeuGlyGlnThrValAla 1190
3695 CAATTTCTTAATAAATAAATAATTCAGGATTCGAAATACAGTATAC 3744
1191 LeuAlaSerGlyThrLeuSerValLeuHisCysGluAlaIleGlyHi 1207
3745 ATTACAAAAGGACAGAGGTCTCAATATATCTGTGTGACCTTATTACCC 3794
1207 sProArgProThrIleSerTrpAlaArgAsnGlyGluValGlnPheS 1224
3795 CAGTGAGGCCACATATACATGACCAAGGATGAACTTGTACAGCCCT 3844
1224 erAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaPro 1240
3845 CAGTAAATAATTTTGGATGGAACCTGGGAAGATACAGATACAGAATCT 3894
1241 ValGluAlaAspValGlyPheTyrThrCysAsnAlaThrAsnAlaLeuGl 1257
3895 ACAAGGAAGAACACAGCATATATGAATGTCTGTAGCTAATCATCTGG 3944
1257 yTyrAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
3945 TTCAGATGTGGAAAGTCTTCTGTGTGTATGACAGGACCTGTCTATCT 3994
1274 ystThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
3995 TGCTGTGTGAAAGAAATATACCAACACGACAGACACACCATCTGTCTGT 4044
1291 AspIleGlySerThrIleLysThrValGlnGlyValAsnValThrIleAs 1307
4045 GTGTTGGAGCATCTGTGGAGGACCTTGGACGAACACGTGCAATCCG 4094
1307 nCysGlnValAlaGlyValProGluAlaGluValThrTrpPheArgAsnL 1324
4095 ATGTCTGTAAAGGTCTCCTCAGCCTAATATACTTGGTTGAAGAGAG 4144
1324 ysSerLysLeu...GlySerProHisHisLeuHisGluGlySerLeuLeu 1339
4145 GAGATCTCTGAGTGGCAATGTTCCTTGTCTTTTCAATGATCCCTGTG 4194
1340 LeuThrAsnValSerSerAspGlnGlyLeuTyrSerCysArgAlaAl 1356
4195 TTGCAGAATGTTTCCCTTGAATAAGAGAACCTACCTCTGCTCCATAGCCAC 4244
1356 aAsnLeuHisGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspP 1373
4245 CAATGTCTTTGAAAGGACGTGGCACATCTGTATCCACTTGTCTGGAAC 4294
1373 roProGlnValProThrGlnLeuGluAspIleArgAla..... 1385
4295 GAAGATGGCCAGAGAGTAGAATGTATTTCTGCAAGGACATAAAAGTAC 4344
1386 LeuLeuAlaAlaThrGlyProAsnLeuProSerValLeuThrSerProLe 1402
4345 ATTCTCCAGCCACCAACACTAGAACCAACAGC..... 4377
1402 uGlyThrGlnLeuValLeuAspProGlyAsnSerAlaLeuLeuGlyCysP 1419

```

4378 .....AATGACCA..... 4386  
 1419 roileysGlyHisProValProAsnIleThrTrpPheHisGlyGln 1435  
 4387 .....ACAGAGAACCCGCCCT..... 4404  
 1436 ProileValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGlyG 1452  
 4404 ..... 4404  
 1452 nIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyGluPheSerC 1469  
 4404 ..... 4404  
 1469 ysLeuAlaGlnAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuVal 1485  
 4404 ..... 4404  
 1486 IleGlnAspTyrTrpTrpSerValAspArgLeuAlaThrCysSerAlaSe 1502  
 4405 ...CAAGAGCCTTTGGGAGCCTGGTAACCTGGTCACATGTTCTGCCAC 4451  
 1502 rCysGlyAsnArgGlyValGlnGlnProArgLeuArgCysLeuLeu...A 1518  
 4452 CTGTGTTCATTGGGAGCCCGCATTCAGAGACCCCACTGTGTGTATGCCCA 4501  
 1518 snSerThrGluValAsnProAlaHisCysAlaGlyLysValArgProAla 1534  
 4502 ATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4551  
 1535 ValGlnProIleAlaCysAsnArgArgAspCysProSerArgTrpMetVa 1551  
 4552 GCTGGGTTCAGGCCTGTAACATCCGGGAGCTGCCAGGAGGTGGTTCAC 4601  
 1551 lThrSerTrpSerAlaCysThrArgSerCysGlyGlyValGlnThrA 1568  
 4602 AAGT 4651  
 1568 rArgValThrCysGlnLysLeuLysAlaSerGlyIleSerThrProVal 1584  
 4652 GGCAGGTGACGTGCAACGGACAAACCAATGGAAGTGTGCAGTGTGTGT 4701  
 1585 SerAsnAspMetCysThrGlnValAlaLysArgProValAspThrGlnAl 1601  
 4702 TCTCCAGAGCATGTGCCCT...AAGACCGGCTCTGGGAGAAACAC 4748  
 1601 aCysAsnGlnLeuCysValGluTTPAlaPheSerSerTrpGlyGlnC 1618  
 4749 ATGTTTGGTATCCATGTTGTAGTGG.....GAACAGGGAACCGGT 4792  
 1618 ysAsnGlyProCysIleGlyProHisLeuAlaValGlnHisArgGlnVal 1634  
 4793 GTCTGGACGTTGATGGCGGTCTGTAGGATGCAGCAGCGTCACACA 4842  
 1635 PheCysGlnThrArgAspGlyIleThrLeuProSerGluGlnCysSerAl 1651  
 4843 GCTTGTCAACACACAGCTCT.....GACTCCCACTGTATGA 4880  
 1651 aLeuProArgProValSerThrGlnAsnCysTrpSerGluAlaCysSerV 1668  
 4881 CAGAAAGAGACCCACCTTAAGAGGAAGTGCATCAGGGCCGTGTGATG 4930  
 1668 aHisTrpArgValSerLeuTrpThrLeuCysThrAlaThrCysGlyAsn 1684  
 4931 TGTGTGSCACACAGCCCTTGAAGCCCTGTACAGCAGCCTGTGGCAGG 4980  
 1685 TyrGlyPheGlnSerArgValGluCysValHisAlaArgThrAsnLy 1701  
 4981 ...GGTTTCCAGTCTCGGAAAGTCACTGTATCCACAAAGAGGTGCAC 5027  
 1701 sAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpG 1718  
 5028 ACCTGTGCCAAGAGACACTGTGTACAGAAAGAAAGAACCAATTCCTGGC 5077

1718 lnArgCysAsnIleThrProCysGluAsnMetMetGluCysArgAspThrThr 1734  
 5078 GGCACGT...CTGGGCCCTCCTGTGATAGAGACTGCACAGACACAACT 5124  
 1735 ArgTyrCysGluLysValLysGlnLeuLysLeuCysGlnLeuSerGlnPh 1751  
 5125 CACTACTGTATGTTTGAACATCTTAATTTGTCTCTAGACCGCTA 5174  
 1751 eLysSerArgCysCysGlyThrCys 1759  
 5175 CAAACAAGGTGTCGCCAGTCATGT 5199

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAD14371

seq\_documentation\_block:

ID AAD14371 standard; cDNA; 4854 BP.

XX AAD14371;

XX DT 01-NOV-2001 (first entry)

XX DE Novel human protein (NHP) cDNA #8.

XX KW Novel human protein; NHP; thrombospondin; gene therapy; cancer;

KW cytostatic; antisense therapy; angiogenesis; biological disorder; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 1..4854

XX FT /\*tag= a

XX FT /product= "Novel human protein"

XX PN WO200161011-A2.

XX PD 23-AUG-2001.

XX PF 15-FEB-2001; 2001WO-US05290.

XX PR 17-FEB-2000; 2000US-0183282.

XX PA (LEXI-) LEXICON GENETICS INC.

XX PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;

PI Sands AT;

XX DR WPI: 2001-514776/56.

XX DR P-PSDB; AAE07870.

XX PT An isolated nucleic acid encoding a thrombospondin useful as a hybridization probe and gene therapy treatments of cancer -

XX PS Claim 4; Page 48-49; 56pp; English.

XX CC The invention relates to novel human polynucleotides encoding proteins that share sequence similarity with animal proteins having thrombospondin repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes, prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and gene trapped human cells. NHPs are thrombospondins useful for treating biological disorders involving angiogenesis, cancer and development and also in pharmacogenomic applications. NHPs are useful as a hybridisation probe for screening libraries, assessing gene expression patterns and also in gene therapy. Proteins having thrombospondin repeats act as receptors, secreted extracellular matrix proteins and proteases. The present sequence is a cDNA encoding novel human protein (NHP).

XX SQ Sequence 4854 BP; 1331 A; 1175 C; 1276 G; 1072 T; 0 other;

alignment\_scores:

Quality: 3762.50

Ratio: 3.194

Length: 1733

Gaps: 31

Percent Similarity: 67.975 Percent Identity: 42.527

alignment\_block:

US-10-044-807-2 x AAD14371 ..

Align seg 1/1 to: AAD14371 from: 1 to: 4854

```
64 LysSerCysGluGlyArgAsnIleArgTyrArgThrCysSerAsnValAs 80
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
94 AGCAATTTGTCAGAGGCGAGACAATCGGTACAGACATGCAGCAATCATGA 143
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
80 pCysProGluAlaIleAspPheArgAlaGlnGlnCysSerAlaHisA 97
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
144 CTGCCCTCCAGATGCAGAGATTTCAGAGCCCGACAGTGTCTCAGCCTACA 193
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
97 snAspValIysHisHisGlyGlnPheTyrGluTyrLeuProValSerAsn 113
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
194 ATGATGTCCAGTATCAGGGCATTTACTATGAATGGCTTCCACGATATAAT 243
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
114 AspProAspAsnProCysSerLeuLysCysGlnAlaLysGlyThrThrLe 130
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
244 GATPCTTGCTGCCCGGTGTGCACCTCAAGTGTGGAATGTCATGCACAAAG 293
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
130 uValValGluLeuAlaProLysValLeuAspGlyThrArgCysTyrThrG 147
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
294 GGTGGTGGAGCTGGCACCTTAAGGTACTGGATGGAATCGTTGCAACACGG 343
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
147 luSerLeuAspMetCysIleSerGlyLeuCysGlnIleValGlyCysAsp 163
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
344 ACTCCTTGGACATGTATCAGTGGCATCTGTCTCAGGCAGTGGGCTCGAT 393
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
164 HisGlnLeuGlySerThrValLysGluAspAsnCysGlyValCysAsnG 180
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
394 CGCAACTGGGAAGCACTGAAGAGGAGCAACTGTGGAGTCTGTGCCGG 443
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
180 yAspGlySerThrCysArgLeuValArgGlyGlnTyrLysSerGlnLeu 197
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
444 CGATGCTCCAGCTCAGGCTGTGAGGGACAAATCAAAAGTCAACGCTTT 493
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
197 erAlaThrIysSerAspThrValValAlaIleProTyrGlySerArg 213
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
494 CTCCTCAAAAAGAGAGAAATGTAATGCTGTCTCTTGGGAAGTCGA 543
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
214 HisIleArgLeuValLysGlyProAspHisLeuTyrLeuGluThrLy 230
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
544 AGTGTGAGAAATTACAGTGAAGAGCACTGCCACCTCTTTATTGAATCAA 593
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
230 sThrLeuGlnGlyThrLysGlyGluAsnSerLeuSerThrGlyThrP 247
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
594 AACACTTCAGGAGCAAGAGGAGAACACAGCTTTAACAGCCCGGGGTCT 643
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
247 heLeuValAspAsnSerSerValAspPheGlnLysPheProAspLysGlu 263
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
644 TTGCTGTAGAAAACACACAGTGAATTTTCAGAGGGGCTCCGAGAGGCCAA 693
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
264 IleLeuArgMetAlaGlyProLeuThrAlaAspPheIleValLysIleAr 280
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
694 ACTTTTAAGATTCCAGGACCTCTGATGGCTGATTTTCATCTTCAAGACCAG 743
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
280 gAsnSerGlySerAlaAspSerThrValGlnPheIlePheTyrGlnProI 297
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
744 GTACACTGCAGCCAAAGACAGCGTGGTTTCAGTTCTTTTACAGCCCA 793
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
297 leIleHisArgTrpArgGluThrAspPhePheProCysSerAlaThrCys 313
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
794 TCAGTTCATCAGTGGACAAACAGTCTTCTTCCCTGCAGTGTGACGTGT 843
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
314 GlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAs 330
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
844 GGAGAGAGTTATCAGTCAATTTCTGTGAATGTGTGCATATCCGCTTGAA 893
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
330 nArgValValAlaAspGlyTyrCysHisTyrTyrProGluAsnIleLysP 347
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

```
894 GAGGTAGTTCTCGACCATATTGTCTACTACTACCTGAAAAATGTA AAC 943
347 roLysProLysLeuGlnGluCysAsnLeuAspProCysProAlaSerAsp 363
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
944 CAAACCAAACTGAGGAATGACGATGATGATCCCTCCCATCAAGTCAT 993
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
364 GlyTyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTr 380
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
994 GGAATTTAAAGAGATAATGCCCTATGACCACTTCCAACTCTTCTCCTCG 1043
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
380 pGluAlaThrProTrpThrAlaCysSerSerCysGlyGlyGlyLeG 397
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1044 GGAACATAATCTTGGACTGCTGATGTTCCGTGCTGTTGGAGGAGGATTC 1093
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
397 lnSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1094 AGACACGGAGCTTGTGTGTAGAGAAATCCATGCATGGAGAGATATTG 1143
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
414 SerValGluGluTrpLysCysMetTyrThrProLysMetProIleAlaGl 430
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1144 CAGGTGCAAGAATGGAAGTGCATGTACGCCCAACCCCAAGGTATATCCA 1193
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
430 nProCysAsnIlePheAspCysProLysTrpLeuAlaGlnGluTrpSerP 447
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1194 AACTTGTAATCTGTTTATTGCCCAAGTGGATTGCCATGGAGTGTCTC 1243
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1244 AGTGACAGTGCATGTGGCCGAGGGTTACGGTACCGGGTTGTCTGTGT 1293
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1294 ATTAACCAACCGCGAGAGCATGTGTGGGGCTGCAATCCCAACTGAAGTT 1343
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
480 oHisIleGluGluCysIleValProThrProCysTyrLysProLysG 497
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1344 ACATATCAAAAGAAATGTGTCTATCCATCCCGTGTATATAAACCAAAAG 1393
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
497 luLysLeuProValGluAlaLysLeuProThrPheLysGlnAlaGlnGlu 513
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1394 AAAAAAGTCCAGTGGAAAGAAAATTCCTTGGCTGAAACCAAGCAAGAA 1443
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1444 CTAGAGAGACCAAGAAATAGCAACAGAGAACCAACTTCATTCAGAAC 1493
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
530 atPpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1494 CTGGTCAGCCTGCAGTACCAGTGTGGGCCAGGTGTGCAGGTCCGCGAG 1543
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
547 alArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAlaAspLeuPro 563
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1544 TGAAGTGCCTGTGCTCCTCAATTCACGACACTGAGACTGAGCTGAGCTGCC 1593
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
564 IleAspGluCysGluGlyProLysProAlaSerGlnArg..... 576
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1594 GAGGAAGAGTGTCAAGGCCCAAGTGCACCAAGCCAGCGGCTGCCTCT 1643
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
577 ....AlaCysTyrAlaGlyProCysSerGlyGluIleProGluPheAsn 592
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1644 GGAAGCATGTGATGAGAGCCCGGCTCCGAGAGACTAGACATCCCTCTCC 1693
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
592 roAspGluThrAspGlyLeuPheGlyLeuGlnAspPheAspGluLeu 608
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1694 CT.....GAGGACACTGAGACGACT 1713
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyGl 625
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1714 TACGACTGGGAGTACCTGGGTTTACCCCTTGACACGACACATGCTTGGG 1763
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
625 yGlyValGlnGluAlaValSerCysLeuAsnLysGlnThrArgGluP 642
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1764 AGGCCATCAAGAACCATAGACAGTGTGCTTACATATCCAGACCCAGACA 1813
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

642	roAlaGluGluAsnLeu	CysValThrSerArgArgProProGlnLeu	658
1814	CAGTCAATGACAGCTTGTGTGATATGGTCTCCACCGTCTCCAGCCATGAGC	1863	...
659	LysSerCysAsnLeu	AspProCysProAlaArgTrpGluLeu	675
1864	CGGCGGTGAACACAGAGCCCTGTCCCCACAGTGGCATGTGGCTCTTG	1913	...
675	pSerProCysSerLeu	ThrCysGlyValGlyLeuGlnThrArgAspValp	692
1914	GGGCGCTGCTCACTACCTGTGGAGTTGGAATTCAGACCCGASATGTGT	1963	...
692	heCysSerHisLeu	LeuSerArgGluMetAsnGluThrValIleLeu	708
1964	ACTGCTGCAC.....CCAGGGAGACCCCTGCCCTCCCT	1998	...
709	AspGluLeuCysArgGlnPro	LysProSerThrValGlnAlaCysAsnAr	725
1999	GAGGAG...TGCCGAGATGAAGACCCCATCGTTTACAGCATGCAATCA	2045	...
725	gPheAsnCysProProAla	TrpTyProAlaGlnTrpGlnProCysSerA	742
2046	GTTTCACTGCCCTCCTGGCTGGCACATTGAAAGATGCGCAGCATGTGCCA	2095	...
742	rgThrCysGlyGlyGlyValGlnLysArgGluValLeuCysLysGlnArg	758	...
2096	GGACTGTGTGGCGGGGAACTCAGAACAGAGAGTCACTCTCGGCAGCTG	2145	...
759	MetAlaAspGlySerPheLeu	GluLeuProGluThrPheCysSerAlaSe	775
2146	CTAAGGATGCGACTTTTGAATCTTCAGATGAATGTGCCAAGACC	2195	...
775	rIlysProAlaCysGlnGlnAlaCysLysLysAspCysProSerGluT	792	...
2196	CAAGGCATCGTCTCACAAGTCTGTGCCAGGACAGACTGCTCCACATT	2245	...
792	rpLeuLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyThr	808	...
2246	TAGCTGTGGGAGACTGTGCGAAGTGTCTGCACTGTGTGTGTGGAACT	2295	...
809	GlnThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerTh	825	...
2296	CAGACAAAGAACAGGTGTGTCAAGGCTGGCAGCCAAAGTGGCGCAT	2345	...
825	rValValAsnSerThrLeuCysProProLeuProPheSerSerIleA	842	...
2346	CCCCCTCAGTCAGATGATGTGCGAGGATCTACCGGTTTCCCTCTGTAA	2395	...
842	rqProCysMetLeuAlaThrCysAlaArg...	ProGlyArgProSerThr	857
2396	GATCTTGGCAGATGCCGTGAGTGCAGTAAATCAATCAGAGATGAAGACA	2445	...
858	Lys.....HisSerProHisIleAlaAlaArgLysValTyrl	871	...
2446	AAACTTGTGTGACAGGTCGCCAGATCCTCAGTGTCCAGAGATCTACAT	2495	...
871	eGlnThrArgArgGlnArgLysLeuHisPheValValIcIlyPheAlat	888	...
2496	TCAGACAAGGAAGAGACGCTATTAACTGACCATTTGGTAGCAGACCT	2545	...
888	yrLeuLeuProLysThrAlaValValLeuArgCysProAlaArgArgVal	904	...
2546	ATTCTGCCAACACATCCGTGTATTAAAGTCCCGCTGCGACGATTC	2595	...
905	ArgLysProLeuIleThrTrpGluLysAspGlyGlnHisLeuLeuSerSe	921	...
2596	CAGAAATCTCTGATCCAGTGGGAGAGGATGGCGTTGCCCTGCAGAACTC	2645	...
921	rThrHisValThrValAlaProPheGlyTyThrLeuLysIleHisArgLeuL	938	...
2646	CAAAAGGCTTGGCATCAGCAAGTCAGGCTACTATAAATCCACGGTCTTG	2695	...

[illegible]



```

3464 AGNTGGAACCTGTTACGCCCTCAGTAAATAATTTGGATGGAACT 3513
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1233 AspSerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheIleThr 1249
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3514 GGGAAAGATACAGATACAGATCTCACAAGAAAGAAACAGGCATATATGA 3563
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1249 rCysAsnAlaThrAsnAlaLeuGlyIleThrAspSerValSerIleAlaValT 1266
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3564 ATGTTCTGTAGCTAATCATCTGTTGGTTCAGATGTGGAAGTCTTCTCTGCG 3613
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1266 hrLeuAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsn 1282
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3614 TGTATGAGAGGACCTGCTCATCTGTTGTTGGAAGAAATATACCAAA 3663
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1283 ThrGluLysProAlaValThrValAspIleGlySerThrIleLysThrVa 1299
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3664 CCAGAGCACAAACCATCTGCTGTTGTTGGTGGAGGCATCTGGGAGGCAGC 3713
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1299 lGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluA 1316
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3714 CMTGGAGCAACAGTGCATCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3763
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1316 laGluValThrTrpPheArgAsnLysSerLysLeu...GlySerProHis 1331
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3764 CTAATATACTTGTTGAGAGGAGGAGTCTCTGAGTGGCAATGTTTCC 3813
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1332 HisLeuHisGluGlySerLeuLeuLeuThrAsnValSerSerAspG1 1348
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3814 TTGCTTTTAAATGGATCCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3863
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1348 nGlyLeuThrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGluS 1365
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3864 AGGAACCTAGCTGTCATAGCCCAATGCTCTTGGAAAGGCACTGGCAA 3913
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1365 erThrGlnLeuLeuIleLeuAspProProGlnValProThrGlnLeuGlu 1381
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3914 CATCTGACTCTCCTGCTGAGAGAGAGATGGCCAGAGATGAAATCGTA 3963
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1382 AspIleArgAla.....LeuLeuAlaAlaThrGlyProAsnLe 1394
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
3964 TTTCTGAGAGGACATAAAAGTAGTATCTCCAGGCAACCAACACTAGAAC 4013
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1394 uProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuAspProG 1411
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4014 CARACAC.....AATGACCCA. 4029
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1411 lYasnSerAlaLeuLeuGlyCysProIleLysGlyHisProValProAsn 1427
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4030 .....ACAGAGAACCCCGCCT... 4047
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1428 lIleThrTrpPheHisGlyGlyGlnProIleValThrAlaThrGlyLeuTh 1444
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4047 ..... 4047
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1444 rHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnLeuSerG 1461
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4047 ..... 4047
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1461 lYcGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAlaGlyVal 1477
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4047 ..... 4047
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1478 LeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSerValAs 1494
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4048 .....CAAGAGCCTTTTGGAGGCGCTGG 4070
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1494 pArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValGlnGlnP 1511
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4071 TAACTGTCACATGTTCTGCCACCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4120
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1511 roArgLeuArgCysLeuLeu...AsnSerThrGluValAsnProAlaHis 1526
      :::::||||| :::::||||| :::::||||| :::::||||| :::::

```

```

4121 AGAGACCCCACTGTGTGATGCCAATGGCAGGAAGTGAGTGAGGCCCTG 4170
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1527 CysAlaGlyLysValArgProAlaValGlnProIleAlaCysAsnArgAr 1543
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4171 TGTGATCACTCCAGACCCACTGGCTGGTGGTTGAGCCCTGTAAACATCCG 4220
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1543 gAspCysProSerArgTrpMetValThrSerTrpSerAlaCysThrArgS 1560
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4221 GGACTGCCCAAGAGGTGGTTTCAACAAGTGTGGTCAACAGTCTCTGTGT 4270
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1560 erCysGlyGlyGlyValGlnThrArgArgValThrCysGlnLysLeuLys 1576
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4271 CTTCCGGTGAAGGATACACACAGTGGCAGGTGAGTGCAGTCTCTGTGT 4320
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1577 AlaSerGlyIleSerThrProValSerAsnAspMetCysThrGlnValAl 1593
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4321 GCCAATGGAACGTGCAGGTGTGTCTCCAAAGAGCATGTGCCCT...AA 4367
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1593 aLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuLysValGluT 1610
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4368 AGACCGGCTCTGGGAAGAAACCATGTTTGGTCACTCCATGTGTTCAGT 4417
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1610 rPalapheSerSerTrpGlyGlnCysAsnGlyProCysIleGlyProHis 1626
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4418 GG.....GAACACAGGAACCGGTGCTGTCAGCTTGCATGGCGGTGCT 4461
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1627 LeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGlyIleTh 1643
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4462 GTGAGGATGAGCAGCGTCACACAGCTTGTCAACACACACAGCTCT.... 4506
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1643 rLeuProSerGluGlnCysSerAlaLeuProArgProValSerThrGlnA 1660
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4507 .....GACTCCACACGTGATGACAGAAAGACACCCACCTTAACAAGGA 4549
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1660 snCysTrpSerGluAlaCysSerValHisTrpArgValSerLeuTrpThr 1676
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4550 ACTGCACATCAGGGGCTGTGATGTGTGTGGTGGCACACAGGCGCTTGAAG 4599
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1677 LeuCysThrAlaThrCysGlyAsnTyrGlyPheGlnSerArgArgValG1 1693
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4600 CCTGTACACAGCGCTGGCAGG...GGTTTCCAGTCTCGGAAAGTCGA 4646
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1693 uCysValHisAlaArgThrAsnLysAlaValProGluHisLeuCysSerT 1710
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4647 CTGTATCCACACAAGAGTTCGAAACCTGTGGCCACAGACACACTGTGTAC 4696
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1710 rpGlyProArgProAlaAsnTrpGlnArgCysAsnIleThrProCysGlu 1726
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4697 AGAAAAGAAACCAATTTCTGCGGCACTGT...CTTGGGCGCTCCTGT 4743
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1727 AsnMetGluCysArgAspThrThrArgTyrCysGluLysValLysGlnLe 1743
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4744 GATAGACCTGCACAGACACACTCACTACTGTATGTTGTAACAATCT 4793
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
1743 uLysLeuCysGlnLeuSerGlnPheLysSerArgCysCysGlyThrCys 1759
      :::::||||| :::::||||| :::::||||| :::::||||| :::::
4794 TAATTGTGTCTCTAGACCGCTACAAACAAAGTGTGTCGACCTCACTGT 4842
      :::::||||| :::::||||| :::::||||| :::::||||| :::::

```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAD14372

seq\_documentation\_block:

ID AAD14372 standard; DNA; 8578 BP.

AC AAD14372;

XX 01-NOV-2001 (first entry)

XX Novel human protein gene open reading frame with flanking sequence.

DE Novel human protein; NHP; thrombospondin; gene therapy; cancer;

XX cytosstatic; antisense therapy; angiogenesis; biological disorder; ds.

OS Homo sapiens.

XX WO200161011-A2.  
 XX 23-AUG-2001.  
 XX 15-FEB-2001; 2001WO-US05290.  
 XX 17-FEB-2000; 2000US-0183282.  
 XX (LEXI-) LEXICON GENETICS INC.  
 XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;  
 PI Sands AT;  
 XX WPI; 2001-514776/56.  
 XX An isolated nucleic acid encoding a thrombospondin useful as a  
 PT hybridization probe and gene therapy treatments of cancer -  
 XX Disclosure; Page 53-56; 56pp; English.  
 XX The invention relates to novel human polynucleotides encoding proteins  
 CC that share sequence similarity with animal proteins having thrombospondin  
 CC repeats. NHPs are expressed in human cell lines, pituitary, lymph nodes,  
 CC prostate, testis, adrenal gland, uterus, foetal kidney, foetal lung and  
 CC gene trapped human cells. NHPs are thrombospondins useful for treating  
 CC biological disorders involving angiogenesis, cancer and development and  
 CC also in pharmacogenomic applications. NHPs are useful as a hybridisation  
 CC probe for screening libraries, assessing gene expression patterns and  
 CC also in gene therapy. Proteins having thrombospondin repeats act as  
 CC receptors, secreted extracellular matrix proteins and proteases. The  
 CC present sequence is novel human protein (NHP) gene containing NHP  
 CC open reading frame with flanking sequences.  
 XX Sequence 8578 BP; 2440 A; 1827 C; 2070 G; 2241 T; 0 other;

alignment\_scores:  
 Quality: 3673.00 Length: 2207  
 Ratio: 3.016 Gaps: 36  
 Percent Similarity: 55.188 Percent Identity: 34.798

alignment\_block:  
 US-10-044-807-2 x AAD14372 ..

Align seg 1/1 to: AAD14372 from: 1 to: 8578

```

22 SerSerArgThrAlaArgSerGluGluAspArgAspGlyLeuTrpAspAl 38
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
368 ACCCTCAAGAACACTCGCTCAGATGAAGACAAAGATGGCAACTGGGATGC 417
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
38 atrpGlyProTrpSerGluCysSerArgThrCysGlyGlyGlyAlaSerT 55
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
418 TTGGGGCGACTGGAGTGCATGCTCCCGACCTGTGGGGGAGGACATCAT 467
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
55 YrSerLeuArgCysLeuSerSerSerSerSerSerSerSerSerSerSer 71
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
468 ATTCTCTCGGAGATGTTTGAAGAGGATTTGAAGGCGAGAGCATT 517
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
72 ArgTrpArgThrCysSerAsnValAspCysProProGluAlaGlyAspPh 88
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
518 CGGTACAGAGACATGCACATCATGCTCCCTCCAGATGCAGAGATTT 567
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
88 eArgAlaGlnCysSerAlaHisAsnAspValLysHisGlyGlnP 105
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
568 CAGAGCCAGCAGTGCCTCAGCCTACATGATGTCCAGTATCAGGGCAT 617
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
105 heTrpGluTrpLeuProValSerAsnAspProAspAsnProCysSerLeu 121
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
618 ACTATGAATGGCTCCACGATATATATGATCTGCTGCCCGGTGCACTC 667
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
122 LysCysGlnAlaLysGlyThrThrLeuValValGluLeuAlaProLysVa 138
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::

```

```

668 AAGTGTCAATCCACAAGGACAAAACCTTGGTGGTGGAGCTGCACCTAAGGT 717
138 lLeuAspGlyThrArgCysTyrThrGluSerLeuAspMetCysIleSerG 155
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
718 ACTGGATGGAACCTCGTTTGCACACGCGACTCCTTGGACATGTGTATCAGT 767
155 lYLeuCysGlnIleValGlyCysAspHisGlnLeuGlySerThrValLys 171
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
768 GCATCTGTACGGCAGTGGCTGCGATCGCACTGGGAACGAATGCCAAG 817
172 GluAspAsnCysGlyValCysAsnGlyAspGlySerThrCysArgLeuVa 188
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
818 GAGACAACTGTGGAGTCTGTGCGCGCATGGCTCCACCTGCAGCCCTTGT 867
188 lArgGlyGlnTyrLysSerGlnLeuSerAlaThrLysSerAspAspThrV 205
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
868 ACGGGGACAATCAAGTCAACAGGTTTCTCTGAAAAAAGAGAGAAAAATG 917
205 alValAlaIleProTyrGlySerArgHisIleArgLeuValLeuLysGly 221
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
918 TAATTGTGTCTCTTGGGAAGTCGAAGTGTGAGATTACAGTGAAGAGA 967
222 ProAspHisLeuTyrLeuGluThrLysThrLeuGlnGlyThrLysGlyG 238
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
968 CTTCCCCACCTCTTATTGAATCAAAAAACACTTCAAGGAACCAAGGAGA 1017
238 uAsnSerLeuSerSerThrGlyThrPheLeuValAspAsnSerSerVala 255
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1018 ACACAGCTTTAACAGCCCGCGCTTTTGTGCTAGAAAACACAACAGTGG 1067
255 spPheGlnLysPheProAspLysGluIleLeuArgMetAlaGlyProLeu 271
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1068 AATTTCAGAGGGCTCCGAGAGGCAAACTTTTAAGATTCCAGGACCTCTG 1117
272 ThrAlaAspPheIleValLysIleArgAsnSerGlySerAlaAspSerTh 288
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1118 ATGGCTGATTTCATCTTCAACACCGAGTACACTGCAGCCCAAGACAGGT 1167
288 rValGlnPheIlePheTyrGlnProIleIleHisArgTrpArgGluThrA 305
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1168 GGTTCAGTTCTCTTTTACCAGCCCATCAGTCATCAGTGGAGACAAACTG 1217
305 spPhePheProCysSerAlaThrCysGlyGlyGlyTyrGlnLeuThrSer 321
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1218 ACTTCTTCCCTGCATGTGAGTGTGGAGGAGGTATCAGCTCAATTTCT 1267
322 AlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspGlnTyrCy 338
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1268 CTTAAATGTGTGGATATCCGCTTGAAGAGGTAGTTCTTGACCATATTG 1317
338 sHisTyrTyrProGluAsnIleLysProLysProLysLeuGlnCysA 355
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1318 TCACCTACTACCTGAAATGTAAAAACCAAAACCAAACTGAAGGAATGCA 1367
355 snLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMetProTyr 371
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1368 GCATGGATCCCTGCCCATCAAGTGTAGTATTAAAGAGATTAATGCCCTAT 1417
372 AspLeuTyrHisProLeuProArg..TrpGluAlaThrProTrpThrAla 387
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1418 GACCATTCCAACTCTTCTCGAGCTGGAACATATATCCTTGGACTGCA 1467
388 CysSerSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerCysVa 404
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1468 TGTTCCTGTCTCTGTGGAGGGGATTCAGAGACGGAGCTTTGTGTGTGT 1517
404 lGluGluAspIleGlnGlyHisValThrSerValGluGluTrpLysCysM 421
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1518 ACAGAAATCCATGATGAGAGATATTTCAGGTGGAGAAATGGAAGTGA 1567
421 etTyrThrProLysMetProIleAlaGlnProCysAsnIlePheAspCys 437
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1568 TGTACGCCACCAAAACCAAGGTTATGCAAACTGTAACTCTGTGTGATGTC 1617

```



3370 TTCCATGTTGTTGGGAGGACTCAGTTGGAGATAATTGAATCATGGGG 3419  
 798 ..... 798  
 3420 GCAGTTCCCTCMTACTGTTCTCGTGGTAAATGAGTCTCAGGATC 3469  
 798 ..... 798  
 3470 TGATGTTTATAAGGGTTCCCTTTTCGCTTGGCTCTCATCTCTCT 3519  
 798 ..... 798  
 3520 GCCTGCTGCCAAGTAAGACGTCCCTTTGGCCCTTCCCTTGTCTTGCAT 3569  
 798 ..... 798  
 3570 GATTCGAGCTTCCTTAGCCAGCTGGAACTGACTTATTAACCTCTTT 3619  
 798 ..... 798  
 3620 CCTTATAACTTACCCAGCTCTTGGGTATGCTTTATTATAACAACATAAGAT 3669  
 798 ..... 798  
 3670 TGGACTAATACAGTAGAGGAATGTAAGTGTGCTTATTTCTCATCCTTC 3719  
 798 ..... 798  
 3720 TTAGTAGCAACTCAATAAATACTCTCTTAAGTCAAAATGTCATTAATAAAT 3769  
 798 ..... 798  
 3770 AACTATCCAAATCTCTGTGGTTTATTTAACTCTCTTTTAACTTAG 3819  
 798 ..... 798  
 3820 AGTGTCTTTCGGGAATTAATCAATGTTTAAATAAATCAACATCAAC 3869  
 798 ..... 798  
 3870 AACTCTAATTTTACTTTAAATCTCTTTTCTAATATCTAATAAAT 3919  
 799 ..... Glu.CysSerThrS 803  
 3920 GCATTAAATTTTAAAGTTGAAAAAATAAATAAATAAATAAATGTTCTGCA 3969  
 803 erCysGlyGluGlyThrGlnThrArgSerAlaIleCysArgLysMetLeu 819  
 3970 GTTGTGTGTTGGAATCCAGAGAAGAGCAGGTGTCTCAAGGCTGGCA 4019  
 820 LysThrGlyLeuSerThrValValAsnSerThrLeuCysProProLeuPr 836  
 4020 GCCAAGGTGGGGCATCCCCCTCAGTGAGATGATGTCAGGGATCACC 4069  
 836 oPheSerSerIleArgProCysMetLeuAlaThrCysAlaArg...P 852  
 4070 AGGTTCCCTCTTTGAATGATCTTGCAGATGCCTGAGTGCAGTAAATCA 4119  
 852 roGlyArgProSerThrLys.....HisSerProHisIleAlaAla 865  
 4120 AATCAGAGATGAGACAAACTTGGTGAGCAGGTCCGAGATCCTCAGT 4169  
 866 AlaArgLysValTyrIleGlnThrArgGlnArgLysLeuHisPheVa 882  
 4170 GTCCAGAGAGTCTACATTCACACAAGGAGGAGAGAGGATTAACCTGAC 4219  
 882 lValGlyGlyPheAlaIleLeuLeuProLysThrAlaValValLeuArgC 899  
 4220 CATTTGGTAGCAGACCTATTGCTGCCCAACACATCCGTGATTATTAAGT 4269  
 899 ysProAlaArgValArgLysProLeuLeuThrTTPGluLysAspGly 915  
 915 ..... 915

4270 GCCCGTCCGACGATCCAGAAATCTGTATCCAGTGGGAGAGGATGGC 4319  
 916 GlnHisLeuIleSerSerThrHisValThrValAlaProPheGlyTyrLe 932  
 4320 CGTTGCTGCAGAACTCCAAAGCGTTGGCATCACCAAGTCAGGCTCACT 4369  
 932 uLysIleHisArgLeuLysProSerAspAlaGlyValTyrThrCysSerA 949  
 4370 AAAAATCCAGGCTTCTGCTGCCCGCCACATCGCGGTGTACGGTGCATTG 4419  
 949 laGlyProAlaArgGluHisPheValIleLeuLeuIleGlyAsnArg 965  
 4420 CAGGCTGTGCAGGAAACAGTGTGTCTCAAGCTCATTTGGTACTGACAAC 4469  
 966 LysLeuValAlaArgProLeuSerProArgSerGluGluValLeuAl 982  
 4470 CGGCTCATCGACGCCCA.....GC 4489  
 982 aGlyArgLysGlyProLysGluAlaLeuGlnThrHisLysHisGlnA 999  
 4490 CCTCAGGGAGCCTATAGGGAATATCTGGGATGGACACAGCGAAGCCA 4539  
 999 snGlyIlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAla 1015  
 4540 ATAGTTTGGGAGTCACATGGCACAAA.....ATCAGGCAAAATGTGGAAT 4583  
 1016 AsnProGlySerArgTyr.....AspAspLeuValSerArg..... 1027  
 4584 AACAAAAATGACCTTTATCTGGATGATGACCACATATAGTACCGCCTTT 4633  
 1028 .....LeuLeuGluGlnGlyTyrProGlyGluLeuLeuAla 1041  
 4634 CTTGAGAGCTCTGTAGGCCACTGCAGCAATTCGAGGAAGCAGCAACT 4683  
 1041 erTrpGluAlaGlnAspSerAlaGluArgAsnThrThrSerGluGluAsp 1057  
 4684 CTTGGAGTTCAAGAATAAGCAGTTTGAAGCAGCAGTTAAACAA..... 4727  
 1058 ProGlyAlaGluGlnValLeuLeuHisLeuProPheThrMetValThrGl 1074  
 4728 ...GGAGCA.....TATAGCATGGATACAGC 4750  
 1074 uGlnArgArgLeuAspAspIleLeuGlyAsnLeuSerGlnProGlu. 1090  
 4751 CCAG.....TTGTAGAGCTGATAAGAACATCAGTCAGCTCATGGAAA 4794  
 1091 .....GluLeuArgAspLeuTyrSerLysHisLeuValAlaGlnLeuAla 1105  
 4795 CCGGAGAGTCCAGCGATCATCTTCGTCCTCCAGCTGATATATCAGTGGTG 4844  
 1106 GlnGluIlePheArgSerHisLeuGluHisGlnAspThrLeuLeuLysPr 1122  
 4845 CCCGAATTAGCCAAAGGCACAGCAACACACATGCAG.....TGGCGGGG 4888  
 1122 oSerGluArgArgThrSerProValThrLeuSerProHisLysHisValS 1139  
 4889 CATCCAGGAAGAGACACCTCT..... 4910  
 1139 erGlyPheSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGly 1155  
 4911 .....GCTGCTCAGCTCAGA.....GGGGAACAGGGAGT 4940  
 1156 GlySerArgArgProHisArgLys..... 1163  
 4941 GTGTCCCAAGCTCGCATGCAAAAAACTCAGGCAAGCTGCATTCAGGCC 4990  
 1164 .....ProThrIleLeuArgLysIleSerAlaAlaGlnGlnLeuSerA 1178  
 4991 GAAGGAGCTGTCTCATGAGGCAA...AGCAACCTCCCTCAATTTCAT 5037  
 1178 laSerGluValVal...ThrHisLeuGlyGlnThrValAlaLeuAlaSer 1193  
 5037 TTAATAAAAAACAATAATTCAGGATTTGGAAATACAGTATACATTACAAAA 5087

```

1194 GlyThrLeuSerValLeuLeuHisCysGluAlaIleGlyHisProArgPr 1210
    |||
5088 AGGACAGAGGATCATCATATCTGTGTACCTTATTACCCCGAGTGGGC 5137
1210 oThrIleSerTrpAlaArgAsnGlyGluValGlnPheSerArgI 1227
    |||
5138 CACATATACATGGACCAAGGATGAACCTTGTATACGCCCTCAGTAAAAA 5187
1227 leLeuLeuGlnProAspSerLeuGlnIleLeuAlaProValGluAla 1243
5188 TAAUUTTTGGTGAACCTGGGAAGATACAGATACAAATCCTACAGGAAA 5237
1244 AspValGlyPheThrCysAsnAlaThrAsnAlaLeuGlyTyraSpSe 1260
5238 GAACAAGGCATATATGATGCTGTGTAGTAAATCATCTGTGTTAGATGT 5287
1260 rValSerIleAlaValThrLeuAlaGlyLysProLeuValLysThrSera 1277
5288 GGAAGTCTCTCTGTGTGTATGAGAGGCACCTGTATCTGTGTCTGTGT 5337
1277 rgMetThrValIleAsnThrGluLysProAlaValThrValAspIleGly 1293
5338 AAAGAAATATCACCAAAACAGACACAAACCATCTGTGTGTGTGTGA 5387
1294 SerThrIleLysThrValGlnGlyValAsnValThrIleAsnCysGlnVa 1310
5388 GGCATCTGTGAGGAGCCCTTGGAGCAACAGTGACAAATCCGATCTCTGT 5437
1310 lAlaGlyValProGluAlaGluValThrTrpPheArgAsnLysSerLysL 1327
5438 AAAAGGTGTCCCTCAGCCTAATATACTTGGTGAAGAGGAGGATCTC 5487
1327 eu...GlySerProHisLeuHisGluGlySerLeuLeuLeuThrAsn 1342
5488 TGAGTGGCAATGTTTCTCTTTCATATGATCCCTGTTGTTCACAAT 5537
1343 ValSerSerSerAspGlnGlyLeuTyThrSerCysArgAlaAlaLeuLeuHi 1359
5538 GTTTCCTTGAAATGAAGCAACTAGCTGTGCATAGCCACCAATGCTCT 5587
1359 sGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspProGlnV 1376
5588 TGGAAAGGCAGTGGCAACATCTGTACTCCACTTGTCTGGAACGAGATGGC 5637
1376 alProThrGlnLeuGluAspIleArgAla.....LeuLeuAla 1388
5638 CAGAGAGTAGAATCGTATTTCTGCAAGGACATAAAAAAGTACATTCCTCAG 5687
1389 AlaThrGlyProAsnLeuProSerValLeuThrSerProLeuGlyThrGI 1405
5688 GCAACCAACTAGAACCAACAGC..... 5711
1405 nLeuValLeuAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysG 1422
5712 .....AATGACCCA.....ACAG 5724
1422 lyHisProValProAsnIleThrTrpPheHisGlyGlnProIleVal 1438
5725 GAGAACCCCGCGCT..... 5738
1439 ThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGI 1455
5738 ..... 5738
1455 nValAlaAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaG 1472
5738 ..... 5738
1472 lnAsnGluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAsp 1488
5739 .....CAAGAG 5744

```

```

1489 TyrTrpTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAs 1505
5745 CTTTTTGGAGCCCTGGTAACCTGGTCACATTTGTTCTGCCACCTGTGGTCA 5794
1505 nArgGlyValGlnGlnProArgLeuArgCysLeuLeu...AsnSerThrG 1521
5795 TTTGGAGAGCCGCATTCAGAGACCCCACTGTGTGTGATGGCCAAATGGCAGG 5844
1521 luValAsnProAlaHisCysAlaGlyLysValArgProAlaValGlnPro 1537
5845 AAGTCAGTGAGGCCCTGTGTGATCATCCTCCAGAGCCCACTGGCTGGGTTT 5894
1538 lIleAlaCysAsnArgArgAspCysProSerArgTrpMetValThrSerTr 1554
5895 GAGCCCTGTACATCCGGAGCTGCCAGGAGGTGGTTTCACAAAGTGTGTG 5944
1554 pSerAlaCysThrArgSerCysGlyGlyValGlnThrArgArgValT 1571
5945 GTCACAGTGCCTCTGTCTTGGGTGAAGGATACCACACTCGGCAGGTGA 5994
1571 hrCysGlnLysLeuLysAlaSerGlyIleSerThrProValSerAsnAsp 1587
5995 CGTGCAAGCGGACAAAAGCCAAATGGAACTGTGCAGTGTGTCTCCAAAG 6044
1588 MetCysThrGlnValAlaLysArgProValAspThrGlnAlaCysAsnGI 1604
6045 GCATGTGCCCT...AAAGACCGGCTCTGGGAAGAAACCATGTTTGG 6091
1604 nGlnLeuCysValGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyP 1621
6092 TCATCCATGTCTTCAGTGG.....GAACACAGGAACCGGTGTCTCTGCAC 6135
1621 roCysIleGlyProHisLeuAlaValGlnHisArgGlnValPheCysGln 1637
6136 GTTCATGGGCCGTGTGTGAGGATGCAGCAGCGTCACACAGCTGTGCAA 6185
1638 ThrArgAspGlyIleThrLeuProSerGluGlnCysSerAlaLeuProAr 1654
6186 CACAACAGCTCT.....GACTCCAACCTGTGTATGACAGAAAGAG 6223
1654 gProValSerThrGlnAsnCysTrpSerGluAlaCysSerValHistTrpA 1671
6224 ACCACCTTAAAGAAGAACTGCATCATCAGGGGCTGTGATGTGTCTGGC 6273
1671 rgValSerLeuTrpThrLeuCysThrAlaThrCysGlyAsnTyrglyPhe 1687
6274 ACACAGGCCCTTGGAGCCCTGTACAGCAGCTGTGGCAGG...GGTTC 6320
1688 GlnSerArgArgValGluCysValHisAlaArgThrAsnLysAlaValPr 1704
6321 CAGTCTCGAAAGTCGACTGTATCCACACAGGAGTTGCCAAACCTGTGGC 6370
1704 oGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpGlnArgCysA 1721
6371 CAAGAGACACTGTGTACAGAAAAAGAACCAATTTCTGGCGGCACCTGT. 6419
1721 snIleThrProCysGluAsnMetGluCysArgAspThrThrArgTyrcys 1737
6420 ..CTTGGGCCCTCTCTGTGTAGAGACTGCACAGACACAACTCACTACTGT 6467
1738 GlulysValLysGlnLeuLysLeuCysGlnLeuSerGlnPheLysSerAr 1754
6468 ATGTTTGTAAACATCTTAATTTGTCTCTAGACCGCTACAAACANAG 6517
1754 gCysCysGlyThrCys 1759
6518 GTGCTGCCAGTCATGT 6533

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ98232

seq\_documentation\_block:

ID AAZ98232 standard; cdna; 3045 bp.

XX

AAZ98232;  
 11-MAY-2000 (first entry)  
 Human signal peptide containing protein HSPB-124 cDNA SEQ ID NO:258.  
 Human; signal peptide-containing protein; HSPB; diagnosis; cancer;  
 inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
 antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;  
 antiasthmatic; gene therapy; cell proliferation; neurological disorder;  
 reproductive disorder; developmental disorder; arteriosclerosis;  
 cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
 asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
 Parkinson's disease; Huntington's disease; ovulatory defect;  
 muscular dystrophy; ss.  
 Homo sapiens.  
 WO200000610-A2.  
 06-JAN-2000.  
 25-JUN-1999; 99WO-US14484.  
 26-JUN-1998; 98US-0090762.  
 31-JUL-1998; 98US-0094983.  
 01-OCT-1998; 98US-0102686.  
 11-DEC-1998; 98US-0112129.  
 (INCY-) INCYTE PHARM INC.  
 Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;  
 Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;  
 Bandman O;  
 WPI; 2000-160673/14.  
 P-PSDB; AAY87347.  
 New human signal peptide-containing proteins useful in treatment,  
 prevention and diagnosis of e.g. cancer, inflammation and  
 cardiovascular disease  
 Claim 9; Page 321-322; 327pp; English.  
 AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the  
 human signal peptide-containing proteins HSPB-1 to HSPB-134. HSPBs have  
 anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,  
 neuroprotective, cardiovascular and antiasthmatic activities, and can  
 be used in gene therapy. HSPBs can be used to treat or prevent disorders  
 associated with decreased activity or function of HSPB. Antagonists of  
 HSPB are used to treat or prevent disorders associated with increased  
 activity or function of HSPB. Such diseases include cell proliferation  
 (including cancer), inflammation, cardiovascular, neurological,  
 reproductive or developmental disorders, (e.g. arteriosclerosis,  
 cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,  
 asthma, Crohn's disease, microbial or other infections, congestive or  
 ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
 diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPB  
 nucleic acids can be used for the recombinant production of HSPB, for  
 detecting HSPB in standard hybridisation and amplification assays (for  
 diagnosis and monitoring), in gene therapy, as antisense, as  
 triplex-forming or ribozyme therapeutics, for detecting related sequences  
 or genetic variations, and for chromosomal mapping. HSPB are also used to  
 raise specific antibodies (Ab) and to screen for agonists and  
 antagonists (potential therapeutic agents). Ab are used to diagnose, or  
 monitor, HSPB-related diseases (in usual immunoassays), as therapeutic  
 antagonists, in competitive drug screens, and for purification of HSPB  
 from natural sources.  
 Sequence 3045 BP; 763 A; 772 C; 768 G; 723 T; 19 other;  
 alignment\_scores:

Quality: 3612.00 Length: 860  
 Ratio: 4.914 Gaps: 11  
 Percent similarity: 85.465 Percent identity: 80.116  
 alignment\_block:  
 US-10-044-807-2 x AAZ98232 ..  
 Align seg 1/1 to: AAZ98232 from: 1 to: 3045  
 1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17  
 130 ATGGAATGCGCGTCCGGCAACTCTGGCACACTGCTCTCTTCTTGGC 179  
 17 apheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgaspG 34  
 180 TTTCCTGCTCTGAGTTCAGGACCGCAGCTCCGAGGAGGAGCGGAGC 229  
 34 lyleuTrpAspAlaThrGlyProThrSerGluCysSerArgThrCysGly 50  
 230 GCCTATGGGATGCTGGGGCCCATGGAGTGAAGCTCAGCACCTGCGGG 279  
 51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG1 67  
 280 GGAGGGGCTCTCTACTCTCTGAGCGCTGCTGAGCAGCAGAGCTGTGA 329  
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84  
 330 AGGAAGAAATATCCGATACAGACATGCAGTAATGTGACTGCCACCGAG 379  
 84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspVallys 100  
 380 AAGCAGGTGATTCGAGCTCAGCAATGCTCAGCTCATATAIATGTCAAG 429  
 101 HisHisGlyGlnPheThrGluTrpLeuProValSerAsnAspProAspAs 117  
 430 CACCATGCCAGTTTATGAATGGCTTCTCTGTCTAATGACCTTGACAA 479  
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134  
 480 CCCATGTTCATCAAGTGCCAGCCAAAGAACCAACCTGCTTGTGAAC 529  
 134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150  
 530 TAGCACCTTAAGCTCTAGATGTCGCGTTCCTATACAGAAATCTTGGAT 579  
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167  
 580 ATGTGCATCAGTGGTTTATGCCAAATTTGCTGCGATCACCAGCTGGG 629  
 167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySert 184  
 630 AAGCACCTCAAGGAAGATAACTGTGGGTCTGCAACGGAGATGGGTCCA 679  
 184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200  
 680 CTGCGCGCTGGTCCAGGGCAGTATAAATCCACGCTCCGCAACCAAA 729  
 201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgle 217  
 730 TCGGATGATACTGTGTTGCAATTCCTATGGAAGTAGACATATTCGCT 779  
 217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234  
 780 TGTCTTAAAGGTCTGTGATCACTTATATCTGGAAACCAAAACCTCCAGG 829  
 234 lYThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250  
 830 GCACCTAAGGTGAACACAGCTCTCAGCTCCACAGGAACCTTCCTTGTGGAC 879  
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267  
 880 AATTCTAGTGTGACTTCCAGAAATTTCCAGAAATTTCCAGACAAAGAGAT 929

```

267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
930 GGTGGACCATCTACAGCAGATTTCTATTGTCAGATTCGTAATCGGGCT 979

284 eAlaAspSerThrValGlnPheIlePheThrGlnProIleIleHisArg 300
|||||
980 CCGCTGACAGTACAGTCCAGTTCATCTCTCTACACCCATCATCCACGA 1029

301 TrpArgGluThrAspPheProCysSerAlaThrCysGlyGlyGlyTy 317
|||||
1030 TGGAGGGAGAGAGGATTTCTTCTCTCTGCTCAGCAAGCTGTGGAGGAGTTA 1079

317 rGlnLeuThrSerAlaGluCysTyAspLeuArgSerAsnArgValVala 334
|||||
1080 TCAGCTGACATCGGTGAGTGTACATCTGTGAGGAGCAACCGTGTGGTGG 1129

334 laAspGlnTyCysHisTyTrpProGluAsnIleLysProLysProLys 350
|||||
1130 CTGACCAATACTGTCACTATTACCCAGAGAACATCAAAACCCAAACCAAG 1179

351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyLysG 367
|||||
1180 CTTTCAAGAGTGCACCTTGATTCCTGTCCAGCCAGTGCAGGATACAAGCA 1229

367 nIleMetProTyAspLeuTyHisProLeuProArgTrpGluAlaThrP 384
|||||
1230 GATCATGCTTATGACCTCTACCATCCCTTCTCGGTGGAGGCCACCC 1279

384 roTrpThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
|||||
1280 CATGGACCGGTGCTCTCTCTCTGTTGGGGGGGCATCCAGACCGGGCA 1329

401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417
|||||
1330 GTTTCCTGTGGAGGAGGAGCATCCAGGGCATGTCACTTCAGTGAAGA 1379

417 utrPLysCysMetTyThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1380 GTGGAAATGCATGTACACCCCTAGATGCCCATCGCGAGCCCTGCAACA 1429

434 lePheAspCysProLysTrpLeuAlaGlnLutrpSerProCysThrVal 450
|||||
1430 TTTTGTAGTCCCTTAAATGGCTGCACAGGAGTGGTCTCTCGTCGACAGTG 1479

451 ThrCysGlyGlnGlyLeuArgTyArgValValLeuCysIleAspHisAr 467
|||||
1480 ACGTGTGGCCAGGGCCCTCAGATACCGTGGTCTCTGCTCGACCATCG 1529

467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1530 AGGAATGCACACAGGAGGCTGTAGCCCAAAAACAAAGCCCCACATAAAG 1579

484 luGluCysIleValProThrProCysTyLysProLysGluLysLeuPro 500
|||||
1580 AGGAATGCATCGTACCCACCTCCCTGCTATAAACCCAAAGAAACTTCCA 1629

501 ValGluAlaLysLeuProTrpPheLysGlnAlaGlnGluLutrpSerAla 517
|||||
1630 GTGAGGCCAAGTGGCATGGTTCACAAAGCTCAAGAGCTAGAAGAAGG 1679

517 yAlaIaValSerGluGluProSerPheIleProGluAlaTrpSerAlaC 534
|||||
1680 AGCTGCNGTGTACAGAGAGCCCTCGTTCATCCAGAGGCCCTGGTCGCCCT 1729

534 yThrValThrCysGlyValGlyThrGlnValArgIleValArgCysGln 550
|||||
1730 GCACAGTCACCTTNNNTGTGGGACCCAGGTGGCAATAGTCAGTGCCAG 1779

551 ValLeuLeuSerPheSerGlnSerValAlaAspLeuProIleAspGluCy 567
|||||
1780 GTGCTCTGTCTTCTCAAGNNNTGGCTGACCTGCCTATTACGAGAGTG 1829

567 sGluGlyProLysProAlaSerGlnArgAlaCysTyAlaGlyProCysS 584

```

```

|||||
1830 TGAAGGGCCCAAGCCAGCATCCCGAGGTGCCNNNNATGCAGGCCCATGCA 1879

584 eGlyGluIleProGluPheAsnProAspGluThrAspGlyLeuPheGly 600
|||||
1880 GGGGGGAAATTCCTGAGTTTCAACCCAGACAGACAGATGGNNNTTGGT 1929

601 GlyLeuGlnAspPheAspGluLeuTyAspTrpGluTyGluGlyPheTh 617
|||||
1930 GGCTGCAGGATTTCCAGCAGCTGTATGACTGGGAGTATGAGGGGTTCCAC 1979

617 rLysCysSerGluSerCysGlyGlyGlyValGlnGluAlaValValSerC 634
|||||
1980 CAANNCTCCGAGTCTGTGGAGGAGGTGCCAGGAGGTGTGCTGAGCT 2029

634 yLeuAsnLysGlnThrArgGluProAlaGluGluAsnLeuCysValTh 650
|||||
2030 GCTTGAACAAACANNCTCGGAGCCCTGTGAGGAGCAACCTGTGCGTGAC 2079

650 rSerArgArgProGlnLeuLysSerCysAsnLeuAspProCysP 667
|||||
2080 CACCGCGGCCCCACAGCTCTGAGTCCCTGCAATTTGGATCCCTGCC 2128

667 roAla..ArgTrpGluIleGlyLys.....TrpSerP 677
|||||
2129 CAGCAAGTCTCTCATCTAGGAAGAAGCAGTATCGACTCAGCATGGAACG 2178

677 roCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValPheCys 693
|||||
2179 CTTGCAACGCTCTTTGTTAGGCAACCAAGAGCCCTGGCTTCTCATCTGCC 2228

694 SerHisLeuLeuSer..... 698
|||||
2229 TGTCAACCACTAGCTCTGTGGCTAGGGGAGGTGTCTGCCCTTATGTT 2278

699 .....ArgGluMetAsnGluThrValIleLeuAlaAspGluLeuCysArgG 714
|||
2279 TCCACATCTGCAAAAGTGAACCTGGTTGTA..... 2306

714 lnProLysProSerThrValGlnAlaCysAsnArgPheAsnCysProPro 730
|||||
2307 .....CCTGATGATCTGAGATCCCATGACTTGTCTCACATGTCCTCATG 2348

731 AlaTrpTyProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyCl 747
|||||
2349 ATTCTTAT.....TTTGTAGGCAGAAAG 2371

747 yVal...GlnLysArgGluValLeuCysLysGlnArgMetAlaAspGlyS 763
|||
2372 CATTAACACAGCTACTCTCTGCTGTGTGCTAATCATCTCTGTAAT.... 2417

763 erPheLeuGluLeuProGluThrPheCysSerAlaSerLysProAlaCys 779
|||||
2418 .....TTCTGTCTCTTATTTGCCATTATT 2444

780 GlnGlnAlaCysLysLysAspCysProSerGluTrpLeuLeuSerAs 796
|||||
2445 GAAAAACATGCAAAAGGCTCTTTTACCAACATTCCTGTGTGTAAACA.. 2492

796 pTrpThrGluCysSerThrSerCys...GlyGluGlyThrGlnThrArgS 812
|||
2493 .....ACACCAAAATGCTCAGGCAGTGGCCGAGGAGTCAGTGC 2529

812 eAlaIleCysArgLysMetLeuLysThrGlyLeuSerThrValValAsn 828
|||||
2530 CTGGGACTTGTCTTAAAACTGCTG..... 2552

829 SerThrLeuCysProProLeuProPhe 837
|||
2553 ...GGACTCGTGGTCCCTAAACCCCTTC 2576

```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA06676

see documentation block:



PR 01-DEC-2000; 2000US-250391P.  
 PR 05-DEC-2000; 2000US-251030P.  
 PR 05-DEC-2000; 2000US-251988P.  
 PR 05-DEC-2000; 2000US-256719P.  
 PR 06-DEC-2000; 2000US-251479P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251899P.  
 PR 08-DEC-2000; 2000US-251989P.  
 PR 08-DEC-2000; 2000US-251990P.  
 PR 11-DEC-2000; 2000US-254097P.  
 PR 05-JAN-2001; 2001US-259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-476161/51.  
 DR P-PSDB; ABB10454.  
 XX  
 PT Isolated nucleic acid molecule encoding an inflammation-associated  
 PT polypeptide is used in preventing, treating or ameliorating a medical  
 PT condition -  
 XX  
 PS Claim 1; SEQ ID NO: 342; 859pp + Sequence Listing; English.  
 XX  
 CC The present invention provides human cDNAs, proteins and related genomic  
 CC DNAs, these can be used in the treatment of neural, immune system,  
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,  
 CC renal and proliferative disorders and inflammation. The present sequence  
 CC is a cDNA of the invention.  
 XX  
 SQ Sequence 2031 BP; 449 A; 610 C; 564 G; 408 T; 0 other;

## alignment\_scores:

Quality: 3423.00 Length: 637  
 Ratio: 5.374 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.843

## alignment\_block:

US-10-044-807-2 x ABA06676 ..  
 Align seg 1/1 to: ABA06676 from: 1 to: 2031  
 1126 ArgThrSerProValThrLeuSerProHisLysHisValSerGlyPhe 1142  
 27 AGGACTTCCCGAGTGACTCTCGCCTCAATAACAGCTGTCTGCTCAG 76  
 1142 rSerSerLeuArgThrSerSerThrGlyAspAlaGlyGlySerArgA 1159  
 77 CAGCTCCCTGGGACCTCTCCACCGGGGACGCCGGGGAGGCTCTCGAA 126  
 1159 rgProHisArgLysProThrIleLeuArgLysIleSerAlaAlaGlnGln 1175  
 127 GGCCACACCGCAGGCCACCATCTCGCAAGATCTCAGGGCCAGCAG 176  
 1176 LeuSerAlaSerGluValValThrHisLeuGlyGlnThrValAlaLeuAl 1192  
 177 CTCTCAGCCTCGGAGGTGTCTCCACCTGAGGAGGAGGAGGAGGAGG 226  
 1192 aSerGlyThrLeuSerValLeuHisCysGluAlaIleGlyHisProA 1209  
 227 CAGCGGACACCTCAGTGTCTCTGACATGTGAGGCTATCGGCCACCCAA 276  
 1209 rgProThrIleSerTrpAlaArgAsnGlyGluGluValGlnPheSerAsp 1225  
 277 GGCCATACCATCAGTGGCCAGGATGGAGAGAGATTCAGTTCAGTGAC 326  
 1226 ArgIleLeuLeuGlnProHisPheSerLeuGlnIleLeuAlaProValG 1242  
 327 AGGATTCCTTACGCCAGATGATTCCTTACAGATCTTGGCACCAGTGA 376

1242 uAlaaspValGlyPheTyrThrCysAsuAlaThrAsnAlaLeuGlyTyrA 1259  
 377 AGCAGATGTGGTGTCTACACTTGCAATGCCACCAATGCTTGGGATACG 426  
 1259 spSerValSerIleAlaValThrLeuAlaGlyLysProLeuValIysThr 1275  
 427 ACTCTGTCTCCATTGCCGTACATTAGCAGGAAAGCCACTAGTGAACAG 476  
 1276 SerArgMetThrValIleAsnThrGluLysProAlaValThrValAsp 1292  
 477 TCACCAATGCAGTGTATCAACAGGAGAGCCGTGAGTCACAGTCGATAT 526  
 1292 eGlySerThrIleLysThrValGlnGlyValAsnValThrIleAsnCysG 1309  
 527 AGGAAGCACCATCAAAACAGTGCAGGGAGTGAATGTGACAAATCAACTGCC 576  
 1309 InValAlaGlyValProGluAlaGluValThrTrippheArgAsnLysSer 1325  
 577 AGGTTGCAGAGTGCCTGAAGCTGAAGTCACCTTGTTCAGGAATAAAGC 626  
 1326 LysLeuGlySerProHisHisLeuHisGluGlySerLeuLeuThrAs 1342  
 627 AAACCTGGGCTCCCGCACCATCTGCACGAGGCTCCTTGTCTGCACAAA 676  
 1342 nValSerSerSerAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuH 1359  
 677 CGTGTCTCTCTCGATCAGGCCCTGACTCTGCAGGGCGGCCAATCTTC 726  
 1359 isGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspProProGln 1375  
 727 ATGGAGAGCTGACTGAGAGCACCCAGCTGTGTATCTAGATCCCCCACA 776  
 1376 ValProThrGlnLeuGluAspIleArgAlaLeuAlaAlaThrGlyPr 1392  
 777 GTCCCCACACAGTTGGAAGACATCAGGGCCTGTCTGCTGCTGCTGAC 826  
 1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409  
 827 GAACCTTCCCTCAGTGTCTGCTCTCTCTGGGAAACACAGCTGGTCTGG 876  
 1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425  
 877 ATCTGGGAATTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 926  
 1426 ProAsnIleThrTrippheHisGlyGlnProIleValThrAlaThrGl 1442  
 927 CCTAATATCATCCTGGTTCATGTGTGTGTCAGCCCAATTCCTGCTCAG 976  
 1442 YLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459  
 977 ACTGAGCATCATCATCTTGGCAGCTGGACAGATCCTTCAAGTTGCAACC 1026  
 1459 euSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAla 1475  
 1027 TTAGCGGTGGGTCTCAAGGGGAATTCAGCTGCTGCTGCTGCTGCTGCTG 1076  
 1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrippSe 1492  
 1077 GGGGTCTCATGTCAGAGGATCTTTAGTATCCAAAGATTTACTGGTGGTC 1126  
 1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509  
 1127 TGTGGACAGACTGGCAACCTGCTCAGCCTCCTGTGTGTGTGTGTGTGTGT 1176  
 1509 InGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsnProAla 1525  
 1177 AGCAGCCCGCTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1226  
 1526 HisCysAlaGlyLysValArgProAlaValGlnProIleAlaCysAsnAr 1542  
 1227 CACTGGCAGGGAAGGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1276  
 1542 gArgAspCysProSerArgTrpMetValThrSerTrpSerAlaCysThrA 1559



PR 13-MAY-1998; 98US-0085338.  
 PR 13-MAY-1998; 98US-0085339.  
 PR 15-MAY-1998; 98US-0085573.  
 PR 15-MAY-1998; 98US-0085579.  
 PR 15-MAY-1998; 98US-0085580.  
 PR 15-MAY-1998; 98US-0085582.  
 PR 15-MAY-1998; 98US-0085689.  
 PR 15-MAY-1998; 98US-0085697.  
 PR 15-MAY-1998; 98US-0085700.  
 PR 15-MAY-1998; 98US-0085704.  
 PR 18-MAY-1998; 98US-0086023.  
 PR 22-MAY-1998; 98US-0086392.  
 PR 22-MAY-1998; 98US-0086414.  
 PR 22-MAY-1998; 98US-0086430.  
 PR 22-MAY-1998; 98US-0086486.  
 PR 28-MAY-1998; 98US-0087098.  
 PR 28-MAY-1998; 98US-0087106.  
 PR 28-MAY-1998; 98US-0087208.  
 PR 30-JUL-1998; 98US-0094651.  
 PR 11-SEP-1998; 98US-0100038.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX  
 DR WPI: 1999-551358/46.  
 DR P-PSDB; AAY41729.  
 XX  
 PT New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders -  
 XX  
 PS Claim 2; Fig 119; 530pp; English.  
 XX  
 CC The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. AA233891 to  
 CC AA234338, and AA41685 to AA41774 represent polynucleotide and  
 CC polypeptide sequence given in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

alignment\_scores:  
 Quality: 2919.00 Length: 530  
 Ratio: 5.528 Gaps: 0  
 Percent Similarity: 99.623 Percent Identity: 99.057

alignment\_block:  
 US-10-044-807-2 x AA234146 ..

Align seg 1/1 to: AA234146 from: 1 to: 1869

1 MetGluCysCysArgAlaThrProGlyThrLeuLeuPheLeuAl 17  
 |||  
 133 ATGGAATGTCGCGTGGGCAACTCTGSCACACTGCTCTTCTTGGC 182

17 aPheLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34  
 |||  
 183 TTTCCTGCTCTCTGAGTTCAGACCGCAGCTCCGAGGAGGACCGGAGC 232

34 lYLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50  
 |||  
 233 GCCTATGGATGCTCTGGGCCCATCGAGTGAATGCTCAGCACCTTCGCGG 282

51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG 67  
 |||  
 1183 CTTCAAGAGTGCAACTTGGATCTCTTCTCCAGCCAGTCAGCGGATACAGCA 1232

283 GGAGGGGCTCTACTCTCTGAGCGCTGCTGAGCAGCAAGAGCTGTGA 332  
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84  
 |||  
 333 AGGAAGAATAATCCGATACAGAACATGCAGTAATGTGGACTGCCACCAG 382  
 84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100  
 |||  
 383 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATAATGATGTCAAG 432  
 101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117  
 |||  
 433 CACCATGGCCAGTTTATGAATGGCTTCTGTGTCTAATGACCCCTGACAA 482  
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134  
 |||  
 483 CCCATGTTCACTCAAGTCCCAAGCCAAAGGAACAAACCTGGTGTGAAC 532  
 134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150  
 |||  
 533 TAGCACCTAAGTCTTAGATGTCGCGTTCATATACAGAACTCTTTGGAT 582  
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG 167  
 |||  
 583 ATGTGCATCAGTGTGTTATGCCAAATTTGTCCTCGCATCACCAGCTGG 632  
 167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySert 184  
 |||  
 633 AAGCACCGTCAAGGAAGATAACTGTGGGCTGCAACGGAGATGGGTCCA 682  
 184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200  
 |||  
 683 CCTCCCGCTGGTCCGAGGCGAGTATAATCCAGCTCTCCGCAACCAAA 732  
 201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217  
 |||  
 733 TCGGATGATACTGTGGTTGCATCTCCCTATGGAAGTAGACATATTCGCT 782  
 217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234  
 |||  
 783 TGTCTTAAAAAGGTCCTGATCCTATATCTGGAACCAAAACCTCCAGG 832  
 234 lYThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250  
 |||  
 833 GGACTAAAGTGAACAGCTCAGCTCCAGGAACTTCTCTTGTGGAC 882  
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267  
 |||  
 883 AATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATAGTGAAT 932  
 267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284  
 |||  
 933 GGCTGGACCACCTCAGCAGATTTCAATGTCGAAGATTCTGTAACCTCGGCT 982  
 284 exAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300  
 |||  
 983 CGSCTCAGAGTACAGTCCAGTCACTTCTATACACCCATCATCCACCGA 1032  
 301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317  
 |||  
 1033 TGGAGGGAGACGATTTCTTCTCTCAGCAACCTGTGGAGGAGTTA 1082  
 317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValA 334  
 |||  
 1083 TCAGCTGACATCGGCTGAGTGTACGATCTGAGGAGCAACCTGTGGTGTG 1132  
 334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350  
 |||  
 1133 CTGACCAATACTGTCTACTATTACCCAGAGAATCAAAACCCCAACCAAG 1182  
 351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367  
 |||  
 1183 CTTCAAGAGTGCAACTTGGATCTCTTCTCCAGCCAGTCAGCGGATACAGCA 1232

02-DEC-1999; 99WO-US28551.  
02-DEC-1999; 99WO-US28555.  
16-DEC-1999; 99WO-US30095.  
30-DEC-1999; 99WO-US31243.  
30-DEC-1999; 99WO-US31274.  
05-JAN-2000; 2000WO-US00219.  
06-JAN-2000; 2000WO-US00277.  
06-JAN-2000; 2000WO-US00376.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;  
XX  
DR WPI; 2000-611443/58.  
DR P-PSDB; AAB44285.  
XX  
PT Novel PRO polypeptides and polynucleotides used in detection methods,  
PT to target bioactive molecules to specific cells, and to modulate  
PT cellular activities -  
XX  
PS Claim 2; Fig 119; 636pp; English.  
XX  
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
CC sequence tag) sequences which encode secreted or transmembrane PRO  
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
CC activity. The polynucleotides and polypeptides can be used for detecting  
CC the presence of PRO polypeptides in samples, for linking bioactive  
CC molecules to cells and for modulating biological activities of cells,  
CC using the polypeptides for specific targeting. The polypeptide targeting  
CC can be used to kill the target cells, e.g. for the treatment of cancers.  
CC The polypeptide pairs provide specific targeting of bioactive molecules  
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in  
CC the isolation of the PRO polynucleotide sequences.  
XX  
SQ Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

alignment\_scores:  
Quality: 2919.00 Length: 530  
Ratio: 5.528 Gaps: 0  
Percent Similarity: 99.623 Percent Identity: 99.057  
alignment\_block:  
US-10-044-807-2 x AAC78532 ..  
Align seg 1/1 to: AAC78532 from: 1 to: 1869  
1 MetGluCysCysArgAlaThrProGlyThrLeuLeuPheLeuAl 17  
133 ATGAAATGCTGCTGGGCACTCTGGCACACTGCTCTCTTTCTGGC 182  
17 aphLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34  
183 TTTCTGCTCTGAGTTCCAGACCGCAGCTCCGAGGAGGACCGGAGC 232  
34 lyleuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50  
233 GCCATGGATGCCITGGGCCCATGGAGTGAATGCTCAGCACCTCCGGG 282  
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG 67  
283 GGAGGGGCTCTCTACTCTCTAGGGCGCTGCTGAGCAGCAGAGCTGTGA 332  
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84  
333 AGAAGAATAATCCGATACAGACATGCAGTATGTGGAGTGCCTCCACAG 382  
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100  
|||||

367 nIleMetProTyrAspIleuTyrHisProLeuProArgTrpGluAlaThrP 384  
1233 GATCATGCTTATGACCTTACCATCCCTTCTCGGTGGAGGCCACCC 1282  
384 roTrpThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400  
1283 CATGACCGCGTCTCTCTGTTGGGGGGGCGATCCAGAGCGGGCA 1332  
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417  
1333 GTTCTCTGTTGGAGGAGACATCCAGGGCGCATGTCACATCAGTGAAGA 1382  
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434  
1383 GTGGAATGATGTACACCCCTAAGATGCCATCGCGCAGCCCTGCACAA 1432  
434 lPheAspCysProLysTrpLeuAlaGlnLutrpSerProCysThrVal 450  
1433 TTTTTCAGTGCCTAAATGGCTGGCACAGGAGTGGTCTCGTGCACAGTG 1482  
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467  
1483 ACATGTGGCGAGGGCTCAGATACCTGCTGGTCTGCTGATCGACCATCG 1532  
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484  
1533 AGAATGTCACACAGGAGCTGTAGCCCAAAACAAAGCCCAATAAAG 1582  
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500  
1583 AGCAATGATCGTACCCACATCCCTGCTATAAACCAAGAGAACTTCCA 1632  
501 ValGluAlaLysLeuProTrpPheLysGlnAlaGlnLeuGluGluG 517  
1633 GTCGAGGCCAAGTTGCCATGGTTCAACCAAGCTCAGAGCTAGAAGAGG 1682  
517 YAlaAlaValSerGluGluProSerPheIleProGluAla 530  
1683 AGCTGTGTGTGAGAGGAGCCCTCGTAAAGTTGTAAGCA 1722  
seq\_name: /SIND1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC78532  
seq\_documentation\_block:  
ID AAC78532 standard; cDNA; 1869 BP.  
XX  
AC AAC78532;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human PRO1071 (UNQ528) nucleotide sequence 58Q ID NO:300.  
XX  
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;  
KW expressed sequence tag; detection; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200053756-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 18-FEB-2000; 2000WO-US04341.  
XX  
PR 08-MAR-1999; 99WO-US05028.  
PR 12-MAR-1999; 99US-0123957.  
PR 29-MAR-1999; 99US-0126773.  
PR 21-APR-1999; 99US-0130232.  
PR 28-APR-1999; 99US-0131445.  
PR 14-MAY-1999; 99US-0134287.  
PR 23-JUN-1999; 99US-0141037.  
PR 26-JUL-1999; 99US-0145698.  
PR 29-OCT-1999; 99US-0162506.  
PR 30-NOV-1999; 99WO-US28313.

```

383 AAGCAGGTGATTTCGAGCTCAGCAATGCTCAGCTCATATAATGATGTCAAG 432
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
433 CACCATGGCCAGTATTTATGATGCTTCCTGTCTAATGACCTGACAA 482
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
483 CCCATGTTCACTCAAGTGCCCAAGGCAAGGCAACCCCTGGTTTGTGAAC 532
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
533 TAGCACCTTAGGTCTTAGATGGTACGGTGTCTATACAGATCTTTGGAT 582
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG 167
583 ATGTGTCATCAGTGGTATATCCAAATGTTGGTGGCATCACCAGCTGGG 632
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
633 AAGCACCGTCAAGGAAGATACCTGTGGGCTGCAACGGAGATGGGTCCA 682
184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
683 CCTGCGCGCTGGTCCGAGGCGAGTATAAATCCAGAGCTCTCCGCAACCAA 732
201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
733 TCGATGATACCTGTGGTGTGACATTCCTATGGAAGTAGACATATTCGCT 782
217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234
783 TGTCTTAAAGGTCCCTGATCATTATATCTGGAACCAAAACCCCTCCAGG 832
234 lyThrLysGlyGluAsnSerLeuSerThrGlyThrPheLeuValAsp 250
833 GGACTAAAGGTGAAACAGCTCAGCTCCACAGGAACCTTCCTTTGGGAC 882
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
883 AATCTTAGTGTGGACATCCAGAAATTTCCAGACAAAGAGATCTGAGAA 932
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
933 GGCTGACCACTCACAGCAGATTTCAATGTCAAGATTCTGTAACCTCGGCT 982
284 exAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
983 CCGCTGACAGTACAGTCCAGTTCATCTTCTATCAACCCATCATCCACGA 1032
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317
1033 TGGAGGGAGACGGATTTCTTCTTCTGCTCAGAACCTGTGGAGAGGTWA 1082
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValVala 334
1083 TCAGCTGACATCGGCTGAGTGTACGATCTGAGGACCAACCCGTGGTGTG 1132
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
1133 CTGACCAATAGTGTCACTATTACCCAGAGAACATCAAAACCCCAACCCAG 1182
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG 367
1183 CTTAGGAGTGGCACTGGATCCTTGTCTCCAGCCAGTGAGGATACAAGCA 1232
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTrpGluAlaThrP 384
1233 GATCATGCTTATGACCTCTACCATCCCTTCTCGGTGGGAGGCCACCC 1282
384 roTrpThrAlaCysSerSerCysGlyGlyGlyGlyGlyGlyGlyGlyGly 400
1283 CATGACCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1332

```

```

401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG 417
1333 GTTTCCTGTGTGGAGGAGGACATCCAGGGCATGCTCACTTCAGTGGAGA 1382
417 uTrpLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
1383 GTGAAATGCATGTACACCCCTAGATGCCATCGCGAGCCCTGCAACA 1432
434 lePheAspCysProLysTrpLeuAlaGlnGluTrpSerProCysThrVal 450
1433 TTTTGTACTGCCCTAAATGGCTGGCACAGAGTGTCTCCGTGCACAGTG 1482
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
1483 ACATGTGCCAGGGCTCAGATACCGTGTGTCTCTGTCATCGACCATCG 1532
467 gGlyMetHisThrGlyGlyCysSerProLysThrLysProHisIleLysG 484
1533 AGGAATGCACACAGSAGCTGTAGCCCAAAACAAAGCCCAATATAAG 1582
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
1583 AGGAATGCATGCTACCCACTCCCTGCTATAAAACCAAGAGAACTTCCA 1632
501 ValGluAlaLysLeuProTrpPheIlyscInAlaGlnGluLeuGluG 517
1633 GTCAGGCCAAAGTTGCCATGTTCAACAAGCTCAAGAGCTAGAAGAAG 1682
517 yAlaAlaValSerGluGluProSerPheIleProGluAla 530
1683 AGCTGCTGTGTACAGAGGAGCCCTCGTAAGTTGTATAAAGCA 1722

```

seq name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS46142

seq\_documentation\_block:

ID AAS46142 standard; cDNA; 1869 bp.

AC AAS46142;

DT 18-DEC-2001 (first entry)

XX Human DNA encoding PRO polypeptide sequence #218.

XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;  
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;  
 KW PCR primer.

XX Homo sapiens.

XX WO200168848-A2.

XX 20-SEP-2001.

XX 28-FEB-2001; 2001WO-US06520.

XX 01-MAR-2000; 2000WO-US05601.

XX 02-MAR-2000; 2000WO-US05841.

XX 03-MAR-2000; 2000US-187202P.

XX 06-MAR-2000; 2000US-186968P.

XX 14-MAR-2000; 2000US-189320P.

XX 14-MAR-2000; 2000US-189328P.

XX 15-MAR-2000; 2000WO-US06884.

XX 21-MAR-2000; 2000US-190828P.

XX 21-MAR-2000; 2000US-191007P.

XX 21-MAR-2000; 2000US-191048P.

XX 28-MAR-2000; 2000US-192655P.

XX 29-MAR-2000; 2000US-193032P.

XX 30-MAR-2000; 2000US-193053P.

XX 30-MAR-2000; 2000WO-US08439.

PR 04-APR-2000; 2000US-194449P.  
 PR 04-APR-2000; 2000US-194647P.  
 PR 11-APR-2000; 2000US-195975P.  
 PR 11-APR-2000; 2000US-196000P.  
 PR 11-APR-2000; 2000US-196187P.  
 PR 11-APR-2000; 2000US-196690P.  
 PR 11-APR-2000; 2000US-196820P.  
 PR 18-APR-2000; 2000US-198121P.  
 PR 18-APR-2000; 2000US-198585P.  
 PR 25-APR-2000; 2000US-199397P.  
 PR 25-APR-2000; 2000US-199550P.  
 PR 25-APR-2000; 2000US-199654P.  
 PR 03-MAY-2000; 2000US-201516P.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 05-JUN-2000; 2000US-209832P.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 22-AUG-2000; 2000US-0644848.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 08-NOV-2000; 2000WO-US30952.  
 PR 01-DEC-2000; 2000WO-US32678.  
 PR 20-DEC-2000; 2000WO-US34956.

(GETH ) GENENTECH INC.

XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;

PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

XX WPI; 2001-602746/68.

DR P-PSDB; AAU29241.

XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
 PT presence of tumours, such as prostate and breast tumours, in mammals and  
 PT to screen for modulators of the compounds -

XX Claim 2; Fig 435; 774pp; English.

XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR  
 CC primers for PRO polypeptides of the invention. The sequences of the  
 CC invention can be used to detect the presence of a tumour in a mammal by  
 CC comparing the level of expression of a PRO polypeptide in a test sample  
 CC of cells from the animal and a control sample of normal cells, whereby a  
 CC higher level of expression in the test sample indicates the presence of a  
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,  
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be  
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human  
 CC blood, when contacted with it. A specific polypeptide can be used to  
 CC stimulate the proliferation or differentiation of chondrocyte cells. The  
 CC PRO proteins can be used to determine the presence of tumours and also  
 CC susceptibility to tumour development, particularly adrenal, lung, colon,  
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
 CC can be used for genetic analysis of individuals with genetic disorders.

XX SQ Sequence 1869 BP; 490 A; 469 C; 480 G; 430 T; 0 other;

#### alignment\_scores:

Quality: 2919.00 Length: 530  
 Ratio: 5.528 Gaps: 0  
 Percent Similarity: 99.623 Percent Identity: 99.057

#### alignment\_block:

US-10-044-807-2 x AAS46142 ..

Align seg 1/1 to: AAS46142 from: 1 to: 1869

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuLeuPheLeuAl 17  
 |||||  
 133 ATGGATGCTGCCGTCGGGCAACTCCGGCACACTGCTCCTCTTCTGGC 182

17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34  
 |||||  
 183 TTTCCTGCTCCTGAGTCCAGGACCGCAGCTCCGAGGAGGACCGGGAGC 232  
 34 LyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50  
 |||||  
 233 GCCTATGGGATGCCCTGGGGCCCATGGAGTGAATGCTCAGCACCTGCGG 282  
 51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG 67  
 |||||  
 283 GGAGGGGCGCTCCTACTCTCTGAGCGCTGCTGAGCAGCAAGAGCTGTGA 332  
 67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProPro 84  
 |||||  
 333 AGSAGAAATATCCGATACAGAACATGCAGTAAATGTGGAGTGGCCACC 382  
 84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100  
 |||||  
 383 AGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATATGATGATCAAG 432  
 101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117  
 |||||  
 433 CACCATGGCCAGTTTATGAATGGCTTCTGTCTTAATGACCTGACAA 482  
 117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGlu 134  
 |||||  
 483 CCATGTTTCACTCAAGTGCACGCAAGCAAGCAACCTGTGTGTGAAC 532  
 134 euAlaProLysValLeuAspGlyThrArgCystYrThrGluSerLeuAsp 150  
 |||||  
 533 TAGCACCCTAAGGCTTAGATGGTACGGTGTGCTATACAGAACTTTGGAT 582  
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG 167  
 |||||  
 583 ATGTGATCATGAGTGTATGCCAAATTTGTTGGCTGCCATCACCAGCTGGG 632  
 167 ySerThrValLysGluAsnAsnCysGlyValCysAsnGlyAspGlySer 184  
 |||||  
 633 AAGCACCGTCAAGGAAGATAACTGTGGGGTCTGCAACGGAGATGGGTCCA 682  
 184 hrCysArgLeuValArgGlyGlnTrpLysSerGlnLeuSerAlaThrLys 200  
 |||||  
 683 CCTGCCGGCTGGTCCGAGGCGAGTATAATCCAGCTCTCCGCAACCAAA 732  
 201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217  
 |||||  
 733 TCGGATCATACTGTGGTGTGCATCTCCCTATGGAAGTAGACATATCGCCT 782  
 217 uValLeuLysGlyProAspHisLeuTyrLeuGluThrLysThrLeuGlnG 234  
 |||||  
 783 TGTCTTAAAGGTCCTGATCATCTTATATCTGGAACCAAAACCTCCAGG 832  
 234 lyThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250  
 |||||  
 833 GGACTAAAGGTGAAACAGATCTCAGCTCCACAGGAACCTTCCCTTGTGGAC 882  
 251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267  
 |||||  
 883 AATTCTAGTGTGGACTTCCAGAAATTTCCAGCAAGAGAGTACTGAGAA 932  
 267 talaglyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284  
 |||||  
 933 GCGTGGACCATCTACAGCAGATTCATTGTCAGATTCGTAACTCGGGCT 982  
 284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300  
 |||||  
 983 CCGCTGACAGTACAGTCCAGTTCATCTTCTATCAACCCCATCATCCACGA 1032  
 301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317  
 |||||  
 1033 TGGAGGGAGACGGATTTCTTCTTCTGTCAGCAACCTGTGGAGGAGTTA 1082  
 317 rGlnLeuThrSerAlaGluCystYrAspLeuArgSerAsnArgValVala 334

```

|||||
1083 TCAGTGCATCGCTGAGTCTAGCATCTGAGGAGCAACCGTGTGGTTG 1132
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1133 CTGACCAATAGTCTCAGTATTTACCCAGAGAAACATCAACACCAACCCCAAG 1182
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG1 367
|||||
1183 CTTGAGAGTGCACCTGGATGCTTGTCTCCAGCCAGTGCAGGATACAGCA 1232
367 nileMetProTyrAspLeuTyrHisProLeuProArgTyrGluAlaThrP 384
|||||
1233 GATCATGCTTATGACCTCTACATACATCCCTTCCTCGGTGGAGGCCACCC 1282
384 roTprThrAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAla 400
|||||
1283 CATGACCGGCTGCTCCTCTGCTGCTGGGGGGGCATCCAGAGCGGGCA 1332
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG1 417
|||||
1333 GTTCTCTGTGTGAGGAGGAGCATCCAGGGGCATGTCTCAGTGAAGA 1382
417 uTprLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1383 GTGAAATGCATGTACACCCCTAAGATGCCATCGCGCAGCCCTGCACCA 1432
434 lePheAspCysProLysTyrLeuAlaGlnGluTyrSerProCysThrVal 450
|||||
1433 TTTTTCGACTGCCCTAAATGGCTGGCAGCAGTGGTCTCGGTGCACAGTG 1482
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1483 ACATGTGGCCAGGGCTCAGATACCGTGTGGTCTCTGCAATCGACCATCG 1532
467 gGlyMetHisThrGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1533 AGGAATGCACAGGAGGCTGTAGCCCAAAACAAAGCCCCACATAAAG 1582
484 luGluCystIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1583 AGGAATGCATCGTACCCACTCCCTGCTATAAACCACCAAGAACTTCCA 1632
501 ValGluAlaLysLeuProThrPheLysGlnAlaGlnGluLeuGluG1 517
|||||
1633 CTCGAGGCCAAGTTCATGGTTCAAACAGCTCAAGAGCTAGAGAAGG 1682
517 yAlaAlaValSerGluGluProSerPheIleProGluAla 530
|||||
1683 AGCTGCTGTGTGAGGAGGCCCTCGTAAAGTGTGTAAGCA 1722
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF63447
seq_documentation_block:
ID AAF63447 standard; cDNA; 1803 BP.
XX AC
XX AAF63447;
XX
XX
XX 14-MAY-2001 (first entry)
XX
XX Human ADAMTS-R1 cDNA sequence.
XX
XX ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
XX tumour cachexia; inflammation;; dermatosparaxis; EDS-VIIC; angiogenesis;
XX Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
XX metastasis; embryogenesis; egg implantation; ADAMTS-R1; ss.
XX
XX Homo sapiens.
XX
XX WO200111074-A2.
XX
XX
XX 15-FEB-2001.
XX

```

```

PF 03-AUG-2000; 2000WO-US21223.
XX
PR 06-AUG-1999; 99US-0369364.
XX
PA (CLEV-) CLEVELAND CLINIC FOUND.
PA (APTE/) APTE S S.
PA (HURS/) HURSKAINEN T L.
PA (HIRO/) HIROHATA S.
XX
XX Apte SS, Hurskainen TL, Hirohata S;
XX WPI; 2001-159978/16.
DR P-PSDB; AAB72290.
XX
PT Murine and human 'A Disintegrin-like And Metalloprotease domain with
PT Thrombospondin type I motifs' proteins and the nucleic acids encoding
PT them, useful for treating e.g. tumours, inflammation and arthritis -
XX
XX Claim 15; Fig 11; 181pp; English.
XX
CC This invention relates to murine and human ADAMTS-N (A disintegrin-like
CC and metalloprotease domain with thrombospondin type I motifs) proteins,
CC designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
CC invention are cDNA sequences encoding the proteins, and antibodies
CC specific for the proteins. The nucleic acid sequences and antibodies
CC used in the prevention, diagnosis and treatment of diseases associated
CC with inappropriate ADAMTS-N expression. Disorders that may be treated
CC using the nucleic acids, proteins and antibodies include, for example
CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
CC in arthritic (both inflammatory and non-inflammatory) disease,
CC angiogenesis, tumour growth and metastases, and they may also be used for
CC controlling embryogenesis and implantation of fertilised eggs. The
CC present sequence represents cDNA encoding human ADAMTS-R1.
XX
SQ Sequence 1803 BP; 475 A; 460 C; 457 G; 411 T; 0 other;

```

```

alignment_scores:
Quality: 2910.00 Length: 559
Ratio: 5.399 Gaps: 1
Percent Similarity: 96.422 Percent Identity: 94.275

alignment_block:
US-10-044-807-2 x AAF63447 ..

Align seg 1/1 to: AAF63447 from: 1 to: 1803

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuPheLeuAl 17
|||||
51 ATGGAATGCTGCGTGGGCAACTCTGGCACACTGCTCTCTTCTGGC 100

17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAsp 34
|||||
101 TTTCTCTGCTCTGAGTCCAGGACCGCACGCTCCGAGGAGGACCGGAGC 150

34 lylLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
|||||
151 GCCTATGGATGCTGGGGCCCATGGAGTGAATGCTCAGCACCTGCGGG 200

51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG1 67
|||||
201 GTTGGGGCGCCCAACTCTCTGAGGCGCTGCTGAGCAGCAAGAGCTGTA 250

67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
251 AGGAAGAAATATCCGATACAGAAACATGCAGTAATGTGGACTGCCACCAG 300

84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100
|||||
301 AAGCAGGTGATTTCGAGCTCAGCAATGCTCAGTCAATATGATGTCAG 350

101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117

```

```

|||||
351 CACCATGCCAGCTTTATGAAATGGCTTCTGTGTCTAATGACCCCTGACAA 400
|||||
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
401 CCCATGTTCACTCAAGTGCACAGCCAAAGGAACAAACCCCTGGTTGTGAAC 450
|||||
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
451 TAGCACCTTAAGGTCCTAGATGTACGGTGTCTATACAGAACTTCTTGAT 500
|||||
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167
|||||
501 ATGTGCATCAGTGGTTATGCCAAATGTTGGCTGCGATCACAGCTGGG 550
|||||
167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
|||||
551 AAGCACCGCTCAAGGAAGATAACTGTGGGTCTGCAACGGAGATGGGTCCA 600
|||||
184 hrCysArgLeuValArgGlyGlnTyrLysSerGlnLeuSerAlaThrLys 200
|||||
601 CTTGCCGGCTGGTCCGAGGCGAGTATAAATCCAGCTCTCCGCAACCAAA 650
|||||
201 SerAspAspThrValValAlaIleProTyrGlySerArgHisIleArgLe 217
|||||
651 TCGGATGATACTGTGCTTCAATTCCTATGGAAGTAGACATATTGCGCCT 700
|||||
217 uValLeuLysGlyProAspHisIleuTyrLeuGluThrLysThrLeuGlnG 234
|||||
701 TGTCTTAAAGGTCCTGATCAGTATATCTTGGAAACCAAAACCCCTCCAGG 750
|||||
234 lyThrLysGlyGluAsnSerLeuSerSerThrGlyThrPheLeuValAsp 250
|||||
751 GGACTAAAGGTGAACACAGTCTCAGCTCCACAGGAATTTCTCTTGTGGAC 800
|||||
251 AsnSerSerValAspPheGlnLysPheProAspLysGluIleLeuArgMe 267
|||||
801 AATTCTAGTGTGGACTTCCAGAAATTTCCACAAAGAGATACTGAGAAT 850
|||||
267 tAlaGlyProLeuThrAlaAspPheIleValLysIleArgAsnSerGlyS 284
|||||
851 GGGTGGACCACTCACAGCAGATTTCATGTCAAGATTCTGTAACCTCGGGCT 900
|||||
284 erAlaAspSerThrValGlnPheIlePheTyrGlnProIleIleHisArg 300
|||||
901 CCGCTGACAGTACAGTCCAGTTCATCTTCATCAACCCATCATCCACCGA 950
|||||
301 TrpArgGluThrAspPhePheProCysSerAlaThrCysGlyGlyGlyTy 317
|||||
951 TGGAGGGAGACGGATTTCTTTCTTCTGTCTCAGCAACCTGTGGAGAGGTTA 1000
|||||
317 rGlnLeuThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValA 334
|||||
1001 TCAGCTGACATCGGCTGAGTGTACGATCTGAGGAGCAACCGTGTGGTTG 1050
|||||
334 laAspGlnTyrCysHisTyrTyrProGluAsnIleLysProLysProLys 350
|||||
1051 CTGACCAATCTGTCACTATTATCCAGAGAAACATCAAAACCCAAACCCAA 1100
|||||
351 LeuGlnGluCysAsnLeuAspProCysProAlaSerAspGlyTyrLysG1 367
|||||
1101 CTTGAGGAGTGCAACTTGGATTCCTTGTCTCCAGCCAGTACGGATACAAGA 1150
|||||
367 nIleMetProTyrAspLeuTyrHisProLeuProArgTyrGluAlaThrP 384
|||||
1151 GATCATGCCCTTATGACCTCTACCACTCCCTTCCCTCGGTGGGAGGCCACCC 1200
|||||
384 roTrpThrAlaCysSerSerCysGlyGlyIleGlnSerArgAla 400
|||||
1201 CATGACCGCGTGTCTCTCTCGTGGGGGGGCGATCCAGAGCCGGGCA 1250
|||||
401 ValSerCysValGluGluAspIleGlnGlyHisValThrSerValGluG1 417
|||||

```

```

1251 GTTTCCTGTGTGGAGGAGACATCCAGGGGCATGTCACCTTCAGTGGGAAGA 1300
|||||
417 utrLysCysMetTyrThrProLysMetProIleAlaGlnProCysAsnI 434
|||||
1301 GTGGAAATGCATGTACACCCCTAAGATGCCATCGCGCAGCCCTGCAACA 1350
|||||
434 lepheaSpCysProLysTyrLeuAlaGlnGluTrpSerProCysThrVal 450
|||||
1351 TTTTTCGACTGCCCTAAATGGCTGGCACAGAGTGGTCTCCGTGCACAGTG 1400
|||||
451 ThrCysGlyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisAr 467
|||||
1401 ACGTGTGGCCAGGGCCTCAGATACCGTGTGGTCTCTCGCATCGACCATCG 1450
|||||
467 gGlyMetHisThrGlyCysSerProLysThrLysProHisIleLysG 484
|||||
1451 AGCAATGCCACACAGAGAGGCTGTAGCCCAAAACAAAGCCCCACATAAAAG 1500
|||||
484 luGluCysIleValProThrProCysTyrLysProLysGluLysLeuPro 500
|||||
1501 AGCAATGCATCGTACCCACCTCCCTCGCTATTAACCCCAAGAGAACTTCCA 1550
|||||
501 ValGluAlaLysLeuProTrpPheLysGlnAlaGlnGluLeuGluG1 517
|||||
1551 GTCGAGGCCAAGTTGCCATGTTCAAAACAAGCTCAAGAGCTAGAAGAAGG 1600
|||||
517 yAlaAlaValSerGluGluProSerPheIleProGluAlaTrp..... 531
|||||
1601 AGCTGTGTGTTCAGAGAGGCCCTCGTAAAGTTGTAAAGACACAGACTGTT 1650
|||||
532 .....SerAlaCysThrValThrCys 538
|||||
1651 TATATTTGAAACCTTTTGTAAAGAAAGACAGTGTCTCAGTGGTTAGCT 1700
|||||
539 GlyValGlyThrGlnValArgIleVal 547
|||||
1701 TTCTAGGTTCTGAACCTAAGTGTAAATC 1727

```





```
247 TCTCCA.....TGG..... 255
382 aThrProThrAlaCysSerSerCysGlyGlyIleGlnSerA 399
256 TCCGAGTGGACCTCTGTCTACGAGCTGTGGCAATGGAATCAGCAGC 304
399 rAlaValSerCysValGluGluAspIleGlnGlyHisValThrSerVal 415
305 GCGGCGCTCCTGCGATAGCCTCACAAACCGATGTGAGGCTCTCTGGTC 354
416 GluGluThrLysCysMetTyrThrProLysMetProIleAlaGlnProCys 432
355 CAG.....ACAGGACCTG 368
432 sAsnIlePheAspCysProLysThrLeuAlaGlnGlu..... 444
369 CCACATTCAGGAGTGTGACAAAGATTTAAACAGAGTGGTGGAGCC 418
445 .....TrpSerProCysThrValThrCysGlyGlnGlyLeuArg 457
419 ACTGGTCCCGTGGTCTATCTGTGTGACATGTGTGTGATGTGTGATC 468
458 TyrArgValValLeuCysIleAspHisArgGlyMetHisThrGly..... 472
469 ACAAGGATCCGGCTCTGCAACTCTCCAGCCGCCAGATGAATGGAAACC 518
473 .....GlyCysSerProLysThrLysProHisIleLysGluGluCysI 487
519 CTGTGAGGCGAGCGGGGAGACCAAGCTGCAAGAAAGAGCCCTGC. 567
487 leValProThrProCysTyrLysProLysGluLysLeuProValGluAla 503
568 .....CCCATCAATGGA 579
504 LysLeu.....ProThrPheLysGlnAlaGlnGluGluGlyAlaAl 519
580 GGCTGGGGTCTTGG..... 594
519 aValSerGluGluProSerPheIleProGluAlaTrpSerAlaCysThrV 536
595 .....TCACCATGGGACATCTGTCTG 616
536 alThrCysGlyValGlyThrGlnValArgIleValArgCysGlnValLeu 552
617 TCACCTGTGGAGGAGGGGTACAGAAAGCT..... 645
553 LeuSerPheSerGlnSerValAlaAspLeuProIleAspGluCysGluG 569
646 .....ActCTCTCTCTGGTCTGACTCTAGAAATGACTGAA...GAGAA 683
569 yProLysProAlaSerGln.....ArgAlaCysTyrAlaGlyP 582
684 CAAGAGTGGCCCAATGAGCTGAGGCGGCCCTCCCTATGCTAT..... 726
582 roCysSerGlyGluIleProGluPheAsnProAspGluThrAspGlyLeu 598
726 ..... 726
599 PheGlyGlyLeuGlnAspPheAspGluLeuTyrAspTrpGluTyrGluG 615
727 CACACAGGAGTTCAG...TACAGAAATACAGGAAATGAGCTGTGATAG 773
615 yPheThrLysCysSerGluSerCysGlyGlyValGlnGluAlaValV 632
774 CTGCACCTGAGTGTAC.....TGTCAGAACTCAGTTA 805
632 alSerCysLeuAsnLysGlnThrArgGluProAlaGluAsnLeuCys 648
806 CCATCTGC.....AAAAGGTGTCTGCCCATCATCATGCC. 843
649 ValThrSerArgArgPro.....ProGlnLeuLeuLysSe 660
```

```
844 TCCAATGCCACAGTTCTCTGATGGAGATGCTCTCTCGCTGTTGGCCAG 893
660 rCysAsnLeuAspProCysProAlaArgTrpGluIleGlyLysTrpSerP 677
894 CGACTCTGCGCAGCATGGCTGGTCTCCATGG.....TCCGAGTGGACCT 937
677 roCysSerLeuThrCysGlyValGlyLeuGlnThrArgAspValPheCys 693
938 CCGTGTCTACGAGCTGTGCAATGAATTCAGCAGCGCGCGCTCTCTGTC 987
694 SerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAlaAspG 710
988 GATAGCTCAACACCATGTGAGGGCTCTCTCGGTC..... 1023
710 uLeuCysArgGlnProLysProSerThrValGlnAlaCysAsn....Argp 726
1024 .....CAGACACGGACCTGCCACATTCAGGAGTGTGACAAAAGAT 1063
726 heAsnCysProProAlaTrp...TyrProAlaGlnTrpGlnProCysSer 741
1064 TTAACAGAGTGGTGGCTGGAGCCACCTGGTCCCGTGTGTCTCTCTGTTCT 1113
742 ArgThrCysGlyGlyValGlnLysArgGluValLeuCysLysGln.. 757
1114 GTGACATGTGTGATGTGTGATCACAAGGATCCGGCTCTCAACTCTCC 1163
758 .....ArgMetalAspGlySerPheLeuGluLeuP 768
1164 CAGCCCCCAGATGAATGGAAACCTGTGAAGC.....GAAGCGC 1204
768 roGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAlaCysLys 784
1205 GGGAGACC.....AAAGCCTGCAAG 1224
785 LysAspAspCysPro.....SerGluTrpLeuLeuSerAs 796
1225 AAAGACGCTGCCCATCAATGAGGCTGGGTCTCTTGG.....TCACC 1268
796 pTrpThrGluCysSerThrSerCysGlyGluGluThrGlnThrArgSerA 813
1269 ATGGACATCTGTCTGTCTACCTGTGGAGGAGGGGTACAGAAAGCTAGTC 1318
813 laIle 814
1319 GTCTC 1323
seq_name: /cgn2_5/ptodata/2/ina/6B_comb.seq:US-08-506-296B-20
seq_documentation_block:
; Sequence 20. Application US/08506296B
; Patent No. 6313265
; GENERAL INFORMATION:
; APPLICANT: Phillips, Greg
; APPLICANT: Cunningham, Bruce A.
; APPLICANT: Crossin, Kathryn L.
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESS: The Scripps Research Institute
; STREET: 10550 NO. 6313265th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: U.S.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,296B
; FILING DATE: 24-JUL-1995
```

```

/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Filling, Thomas
/ REGISTRATION NUMBER: 34,163
/ REFERENCE/DOCKET NUMBER: TSRI 488.0
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 554-2937
/ TELEFAX: (619) 554-6312
/ INFORMATION FOR SEQ ID NO: 20:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3783 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..3783
/ US-08-506-296B-20

```

## alignment\_scores:

```

Quality: 317.50 Length: 1121
Ratio: 0.692 Gaps: 54
Percent Similarity: 40.946 Percent Identity: 20.607

```

## alignment\_block:

```

US-10-044-807-2 x US-08-506-296B-20 ..
Align seg 1/1 to: US-08-506-296B-20 from: 1 to: 3783

834 ProLeuProPheSerSerSileArg.....ProCysMe 845
|||||
45 CCCCTGCTGCTACAGATCCAGACGAATATAAGGACACCATGCG 94
|||||
845 tLeuAlaThrCysAlaArgProGlyArgProSer.ThrLysHisSerPro 861
|||||
95 TAGAGCCACCTGT.....CATCAGGAACAGATCTCCA 126
|||||
862 HisIleAlaAlaAlaArgLysValTyrIleGlnThrArgGlnArgLys 878
|||||
127 .....CGCG 131
|||||
878 sLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLysThrAlav 895
|||||
132 CCTG.....GTTGCTTCCCAACAGATGACA 157
|||||
895 alValLeuArgCysProAlaArgArgValArgLysProLeuIleThrTrp 911
|||||
158 TAGCCTGAATGTGAAGCCAGAGGAGAGCCCAAGTGGAGTTCCGCTGG 207
|||||
912 GluLysAspGlyGlnHisLeuIleSerSerThrHisValThrVal..... 926
|||||
208 ACAGAAAGATGGCATCCACTCAACACCAAGAGAAATGGGTGTAGTGGT 257
|||||
927 .....AlaProPhe...GlyTyrLeuLysIle.....HisArgL 937
|||||
258 GCATGAGGACACCTATCTGGCTCTTCCATCCAGGAGCAACACAGCT 307
|||||
937 euLysProSerAspAlaGlyValTyrThrCysSerAla..... 949
|||||
308 TTGCCAGAGGTTTCAGGCACTATCTGGTGTATGCCAATAAGCTA 357
|||||
950 GlyProAlaArgGluHisPheValIleLysLeuIleClyGlyAsnArgly 966
|||||
358 GGAACCTGCCATGCGCATGATC.....CA 383
|||||
966 sLeuValAlaAlaArgProLeuSerProArgSerGluGluValLeuAlag 983
|||||
384 GCTCGTGGCC..... 393

```

```

983 lYArgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAsn 999
|||||
394 .....GAGSGTCCCCCAAG..... 408
|||||
1000 GlyIlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAlas 1016
|||||
408 ..... 408
|||||
1016 nProGlySerArgTyrAspAspLeuValSerArgLeuLeuGlnGlyG 1033
|||||
408 ..... 408
|||||
1033 lYTrpProGlyGluLeuLeuAlaSerTrpGluAlaGlnAspSerAlaGlu 1049
|||||
409 ..TGCCCGAAGAGACTGTAAACCTGTGAAGTGGAGGAGGAGATCA 456
|||||
1050 ArgAsnThrThrSerSerGluGluAspProGlyAlaGluGlnValLeuHi 1066
|||||
457 GTAGTTCTGCTTGCACCCCTCCACCCAGTCA..... 489
|||||
1066 sLeuProPheThrMetValThrGluGlnArgArgLeuAspIleLeuG 1083
|||||
490 ..GCCCCACCTAGGATCTACTGGATGACACAGCAAGATTTTCGACATC 534
|||||
1083 lYAsnLeuSerGlnGlnProGluGluLeuArg.....Asp 1094
|||||
535 .....AACACAGATGAGCGGGTGTCCATGGGCCCAAGTGGAGAC 573
|||||
1095 LeuTyrSerLysHisLeuValAla.....GlnLeuAlaGlnGluIlePh 1109
|||||
574 CTATATTGTGCCAATGTCTTACCTCAGACAATCATTCAGACTACATCTG 623
|||||
1109 eArgSerHisLeuGluHisGlnAspThrLeuLeu.....LysProSerG 1124
|||||
624 CAAATGCCCACTTCCCTGTACCGGACCATCATTCAAAGGAACCTATTG 673
|||||
1124 luArgArgThrSerProValThrLeuSerProHisLysHisValSerGly 1140
|||||
674 ACCTCCGGGTCAAGCC..... 690
|||||
1141 PheSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGlyGlyse 1157
|||||
691 .....ACCAACAGCATGATTGAC..... 708
|||||
1157 rArgArgProHisArgLysProThrIleLeuArgLysLysSerAlaAlag 1174
|||||
709 .....CGGAAGCCACGCTGCTCTTTCCCAACAAAC..... 738
|||||
1174 lnGlnLeuSerAlaSerGluValValThrHisLeuGlyGlnThrValAla 1190
|||||
739 .....TCCAGCAGCGCGCTGTAGCCTTGCAGGGCCAG..... 771
|||||
1191 LeuAlaSerGlyThrLeuSerValLeuLeuHisCysGluAlaIleGlyHI 1207
|||||
772 .....TCATTGATCTGGAGTGCAATTCCTGCTAGGCGATT 803
|||||
1207 sProArgProThrIleSerTrpAlaArgAsnGlyGluGluValGlnPheS 1224
|||||
804 CCCTACACCCCATCAAGTGGCTGACCCCAAGTACCCCAATGCCA...A 850
|||||
1224 erAspArgIleLeuLeuGln...ProAspAspSerLeuGlnIleLeuAla 1239
|||||
851 CAGACCGTGTATCTACCAAAACCAACAGACCCCTGCAACTACTCAAT 900
|||||
1240 ProValGluAlaAspValGlyPheTyrThrCysAsnAlaThrAsnAlaLe 1256
|||||
901 GTGGGCGAGAGACGACGATGGCGATATACCTGCTTCTGAGAACTCGCT 950
|||||
1256 uGlyTyrAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuV 1273
|||||
951 GGCACATGCCCGCATGCTTACTGTTACTGTGGAGCTGCCCATATT 1000
|||||
1273 allYThrSerArgMetThrValIleAsnThrGluLysProAlaValThr 1289

```

```

1001 GG.....CTGCAGAGGCC..... 1014
1290 ValAspIleGlySerThrIleValGlnGlyValAsnValThrIl 1306
1015 .....CAGAGCCATTGTATGCTCCAGGAGAGACTGCCCGCT 1052
1306 eAsnCysGlnValAlaGlyValProGlnAlaGluValThrTrpPheArgA 1323
1053 AGACTCCCAAGTCCAGGAGGCCCAACACAGAGATCACTTGGAGATCA 1102
1323 snIysSerIysLeu.....GlySerProHisHisLeuHis 1334
1103 ACGGAATGCTATGGAGACGGTGAACAAGCAGACCAAGTACCGGATTGAG 1152
1335 GluGlySerLeuLeuLeuThrAsnValSerSerAspGlnGlyLeuTy 1351
1153 CAGGGTCTCTGATCTGTAGTACGTCGACCACTGACACAATGGTGAC 1202
1351 rSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGluSerThrGlnL 1368
1203 CCAGTGTGAAGCCCGCAACAGCAGATGGCTCTGCTAGCAATGCGCTACA 1252
1368 euLeuLeuLeuAspProProGlnValProThrGlnLeuGluAspIleArg 1384
1253 TTTATGTTGTC.....CAGCTGCCAGCCAGG..... 1278
1385 AlaLeuLeuAlaAlaThrGlyProAsnLeuProSerValLeuThrSerPr 1401
1279 .....ATCTCACAAAAGA 1292
1401 oLeuGlyThrGlnLeuValLeuAspProGlyAsnSerAlaLeuLeuGlyC 1418
1293 CAATCAGACATACATGCACTTGAG...GCAGTACTGCTTACTTGTGCT 1339
1418 ysProIleIysGlyHisProValProAsnIleThrTrpPheHisGlyGly 1434
1340 GCAAGGCTTTGGAGCTTCTGTTCCAGTGTCCAGTGGCTG.....GAT 1383
1435 GlnProIleValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGl 1451
1384 GAGAAGAACCCAGTGTCTGAGGATGAACGATTTTCCCTATGCCAA 1433
1451 yGlnIleLeuGlnValAlaAsnLeuSerGlyGlySerGlnGlyGluPheS 1468
1434 TGGAACCTGAGCATCAGAGACCTCCAGGCCAATGACACTGGAGCTATT 1483
1468 erCysLeuAlaGlnAsnGlu.....AlaGlyVal 1477
1484 TCTGCCAGGCTGCCAATGACCAGAACCAATGTGACCATTTTGGCTAACCTA 1533
1478 LeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSerValas 1494
1534 CAGGTTAAAGAGCAACCCAGATCACACAGGGGCCCGGACCGCAATGA 1583
1494 pArgLeuAla.....ThrCysSerAlaSerCys..... 1503
1564 GAAGAAGGTGCAAGGGTGACATTACGTTGCCAGGCTCTTTGACCCCT 1633
1503 ..... 1503
1634 CTTTGCAGGCCAGCATCATCTGGCGTGAGATGGGAGAGACTACAGGAA 1683
1504 .....GlyAsnArgGlyValGl 1509
1684 CGTGGGACAGTACAAGTATTTCATAGAAGATGGAAACTAGTCAATCCA 1733
1509 n.....GlnProArgLeuArgCysLeuLeuAsnSerTr 1520
1734 GAGCCTGGACTACAGTACGAGGGCACTACAGTTGTGTGGCC...AGCA 1780
1520 hrGluValAsnProAlaHisCysAlaGlyLysValArgProAlaValGln 1536

```

```

1781 CTGAACCTGGATGAGGTGGAGAGCAGGCCACAGCTCTTAGTGGTGGGAGC 1830
1537 Pro..IleAlaCysAsnArgArgAspCysProSer..... 1547
1831 CTGGGCCAGTGCCTCACCTGGAGCTGTCGACGCCACCTGCTGAAGCA 1880
1547 ..... 1547
1881 GAGCCAGGTGCACCTTGTCTGGAGCCCTGCTGAAGACCAACAACCTCTCCA 1930
1548 .....ArgTrpMetValThrSe 1553
1931 TTGAGAGATGATGACATTGAATTTGAGACACAGAAATGGCTCTTCGAGAAA 1980
1553 r.....TrpSerAlaCysThrArg.....S 1560
1981 TGGTTCACTGTGGCAAGGTGCCAGGAATCAGACCTCTACTACCCCTCAA 2030
1560 erCysGlyGlyGlyValGlnThrArg..ArgValThrCysGlnLysLeuL 1576
2031 GCTGTCCCTTATGTCCACTACACCTTTCCGGGTCACTGCCAATTAACAAAT 2080
1576 ysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGlnVal 1592
2081 ATGTCCTGGAGAACCCAGCCCTGTCTCTGAGAGTGTGTCACACCTCAG 2130
1593 Ala.....LysArgProValAspThrGlnAlaCysAsnGlnGln.. 1605
2131 GCAGCCCCAGAGAGAAGACCTGTGGATGTGAGAGGGGAAGGAATGAGAC 2180
1606 .....LeuCysValGluTrpAlaPheSerSerTrpClyGlnCysAsnG 1620
2181 CAACAATATGTCATCATCATGGAAGCCCTTCGGTGGATGATGGAATG 2230
1620 lyProCysIle..... 1623
2231 CCCCCAGATTTCAGTACGCTGACAGTGGCTGCCACAGGGCAAGCAGAG 2280
1624 .....GlyProHisLeuAlaValGlnHisAr 1632
2281 ACCTGGAGAAACAGACCGTGGAGCACCTTTCTGTGTGTCTTAACAC 2330
1632 gLlnValPhe..CysGlnThrArgaspGlyIleThrLeuProSerGluGln 1648
2331 TTCACATTTTGTGCTTATGAGATCAAAGTCCAGCAGTGAACACACAGG 2380
1649 CysSerAlaLeu...ProArg...ProValSerThrGlnAsnCysTrpSe 1663
2381 GCAAGGCCCTGAGCCCGAGCTCACCATTGGCTATTGAGGGGAAGACTAC 2430
1663 rGluAlaCysSerValHisTrpArgValSer..... 1673
2431 CCCCAGGTGAGCCCTGAACTTGAGACATCACAAATCTTCAACTCAAGTAC 2480
1674 .....LeuTrpThr.....Leu 1677
2481 TGTGCTTGTGAGTGGAGCGCTGTGACTTGGCCAGCTTAAGGCCACC 2530
1678 CysThrAlaThrCysGlyAsnTyrGlyPheGlnSerArgValGluCy 1694
2531 TCARGGATACAATGTAACATACTGTTGGAGGGCAGCAGAGAAAGCAC 2580
1694 sValHisAlaArgThrAsnLysAla..... 1702
2581 AGCAAGAGGCATATCCACAAAAGCCACATAGTGGTACCTGCAAAATACCAC 2630
1703 ..ValProGluHis.....LeuCysSerTrpGly 1711
2631 CAGTCCCATCCTCAGTGGTTTGGCCCTTACAGCTCTTACCATGTGGAGG 2680
1712 ProArgPro.....AlaAsnTrp...GlnArgCysAsnIleThrPr 1724
2681 TACAGGCTTTAAATGGCGGGCTTGGGGCTTGGAGTGAATGGACCTTC 2730

```

1757 Glvthr 1758

documentation block:

APPLICANT: Lennon

**TITLE OF INVENTION:** the Nucleotide Sequence

NUMBER OF SEQUENCES: 39

ADDRESSEE: Fay, Sharpe, Beal  
ADDRESSEE: Minnich & McKee

STREET: Suite 700

STATE: Ohio

COMPILER READABLE  
TIF: 44114-2518

MEDIUM TYPE: sto

## OPERATING SYSTEM:

## CURRENT APPLICATION

FILED DATE: NO.  
CLASSIFICATION:

APPLICATION NUMBER

ATTORNEY/AGENT INFO

REGISTRATION NUMBER

TELEPHONE: (216)

TELEX: (216) 980

SEQUENCE CHARACTERISTICS

LIFE: NUCLEIC AC  
STRANDEDNESS. si

MOLECULE TYPE: CDN

ANTI-SENSE: no

ORGANISM: Homo Sapiens

LIBRARY: Stratford

## PUBLICATION INFORMATION

1341 rAsnValSerSerSerAspGlnGluLeuThrGlyLeuThrSerCysArgAlaIaAsnL 1358  
||||| : ||||| : : : : : : : : : : : : : :  
1176 CAAGGTGCACCCAGTGCACAATGGTGACCAATGGTCAGCGGCCGACACC 1225  
eUHisGlyGluLeuThrGluSerThrGlnLeuLeuLeuLeuLeuAspPro 1374  
||||| || : : : : : : : : : : : : : :  
1226 GGCACCGGGCTTTGGCTGGCCAATGCCFACATCTACGTTGTC..... 1266  
1375 GlnValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaIaThrGl 1391  
||||| : : : : : : : : : : : : : :  
1267 CAGCTGCCACGCAAG..... 1281  
1391 yProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValL 1408  
||||| : : : : : : : : : : : : : :  
1282 .....ATCCTGACTGGGACAATCAGACGATACATGGGCTG 1315  
1408 euAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisPro 1424  
: : : : : ||||| ||||| : : : : :  
1316 TCCRAG...GGCAGCAGCTGCCTACCTCTGTGCAAGGCCCTTCGGAGCGCT 1364  
1425 ValProAsnIleThrTrpPhe...HisGlyGlyGlnProIleValThrAl 1440  
||||| : : : : : : : : : : : : : :  
1363 GTGCCAGTGTTCAGTGGCTGGACGAGGATGGACAACAGTCTTCAGGA 1412  
1440 aThrGlyLeuThrHisIleLeuAlaAlaGlyGlnIleLeuGlnValA 1457  
:  
1413 CGAACGGCTTTCOCCTAT.....GCCAAATGGGACCTGGGCGATTTC 1453  
1457 laAsnLeuSerGlySerGlnGlyGluPheSerCysLeuAlaGlnAsn 1473  
:  
1454 GAGACCTCCAGGCCAATCACACCGGACGCTACTTCTGCCTGCCTGCCAAT 1503  
1474 GluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTr 1490  
:  
1504 GACCAAAACAATGTTACCATCATGGCTACCTAGAAGTTAAAGAT..... 1548  
1490 pTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArg 1507  
.....GCAACTCAGA 1558  
1507 lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsn 1523  
: : : : : ||||| : : : : : : : : : : : : : :  
1559 TCACTCAGGGCCCCGCGACGACAAATCGAAGAAAAGGTTCAGGGGTGACC 1608  
1524 ProAlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCy 1540  
||| :  
1609 ...TTCAGTGCACGGCTCCTTTGACCCCTCTTCAGCCCGACGATCAC 1555  
:  
1540 sAsnArgArgAspCys..... 1545  
||| :  
1656 CTGGCGTGGGACGGTCGAGACCTCCAGGAGCTTGGGGACAGTGACAAGT 1705  
1546 .....ProSerArg.TrpMetValThrSerTrpSerAlaCysTh 1558  
: : : : : ||||| : : : : : : : : : : : : : :  
1706 ACTTCATAGAGATGGGCGCTGGTTCATCCACGCTGGACTACAGCGAC 1755  
rArgSerCysGlyGlyValGlnThrArgArgValThrCysGlnLysL 1575  
||||| :  
1756 CAGG..... 1759  
1575 euLysAlaSerGlyLeuSerThrProValSerAsnAspMetCysThrGln 1591  
||||| : : : : : ||||| ||||| : : : : : : : : : : : : : :  
1760 .....GCAACTACAGCTGGTGGCCAGTACCGAACTGGATGGTGGAGA 1804  
1592 ValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysVa 1608  
||||| :  
1805 GTAGGGCCACG..... 1816  
1608 lGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCys 1622  
1817 ..TCITGGTGGTGGGAGCGCTGGCGCGGTGCCACGGCTGGTGC 1858

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-427-497E-1  
seq\_documentation\_block:  
; Sequence 1, Application US/08427497E  
; Patent No. 5969124  
; GENERAL INFORMATION:  
; APPLICANT: Lemmon, Vance  
; TITLE OF INVENTION: A Method for Characterizing the  
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and  
; Patent No. 5969124  
; TITLE OF INVENTION: the Nucleotide Sequence  
; TITLE OF INVENTION: Characterized Thereby  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
; ADDRESSEE: Minnich & McKee  
; STREET: 1100 Superior Avenue  
; STREET: Suite 700  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 44114-2518  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720 kb  
; MEDIUM TYPE: storable  
; COMPUTER: Compaq Prolinea 5100e  
; OPERATING SYSTEM: DOS 5.0  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: 07/904,991  
; FILING DATE: June 26, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Minnich, Richard J.  
; REGISTRATION NUMBER: 24,175  
; REFERENCE/DOCKET NUMBER: CWR 2 149-3-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (216) 861-5582  
; TELEFAX: (216) 241-1666  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3774  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: nucleic acids  
; HYPOTHETICAL: irrelevant  
; ANTI-SENSE: no  
; ORIGINAL SOURCE:  
; ORGANISM: Homo Sapiens  
; INDIVIDUAL ISOLATE: 17-18 week fetus  
; IMMEDIATE SOURCE:  
; LIBRARY: Stratagene cDNA Library 936206  
; CLONE: synthesis of 4 clones  
; PUBLICATION INFORMATION:  
; AUTHORS: Hlavin, Mary Louise  
; AUTHORS: Lemmon, Vance  
; TITLE: Molecular structure and functional  
; TITLE: testing of human L1CAM: an  
; TITLE: interspecies comparison.  
; JOURNAL: GENOMICS  
; VOLUME: 11  
; ISSUE:  
; PAGES: 416-423  
; DATE: 1991  
; RELEVANT RESIDUES IN SEQ ID NO: 1 to 3774

alignment\_scores:  
Quality: 312.00 Length: 815  
Ratio: 0.904 Gaps: 30  
Percent Similarity: 42.331 Percent Identity: 20.368  
alignment\_block:  
US-10-044-807-2 x US-08-427-497E-1 ..  
Align seg 1/1 to: US-08-427-497E-1 from: 1 to: 3774  
843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysH1 859  
|||||:|||||  
46 CCTGCTGCTTATCCAGATC.....CCGAGGAATATCA 80  
859 sSerProHisIleAlaAlaAlaArgLysValTyrIleGlnThrArgArg 876  
:|||||:|||||  
81 AGGACACCATGTGATGGAGCCACCTGTCATCACGGAACAGTCTCCACGGC 130  
:|||||:|||||  
876 InArgLysLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLys 892  
:|||||:|||||  
131 GC.....CTGGTTGCTTCCCCACA 150  
893 ThrAlaValLeuArgCysProAlaArgValArgLysProLeuI1 909  
:|||||:|||||  
151 GATGACATCAGCTCAAGTGTGAGCCAGTGCAGCGGAGTGCAGTT 200  
909 eThrTrpGluLysaspGlyGlnHisLeuIleSerSerThrHisValThrV 926  
:|||||:|||||  
201 CGCTGGACGAGGATGTGCTCACTTCAAAACCCCAAGGAGAGTGGGTG 250  
:|||||:|||||  
926 alAlaProPheGlyTyrLeuLysIleHisArgLeuLysProSerAspAla 942  
:|||||:|||||  
251 TGACCGGTGAC.....CAGTGGCCCTCACT 276  
943 GlyValTyrThrCysSerAlaGlyProAlaArgGluHisPheValIleL 959  
:|||||:|||||  
277 GGCTCCTTCACCATCAGCGCAAC.....AACACCAACTTGTCTCAGAG 320  
959 sLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeuSerProArg 976  
:|||||:|||||  
321 GTTCCAGGGCATCTACCGCTGCTTTGCCAGCAATAAGCTGGCACCAGCA 370  
976 erGluGluGluValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeu 992  
:|||||:|||||  
371 TGTCCTAGATCGGCTCATGCCGAGGTGCCCAAG.....411  
993 GlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGl 1009  
411 ..... 411  
1009 uLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAspLeuVal 1026  
411 ..... 411  
1026 erArgLeuLeuGluGlnGlyTyrProGlyGluLeuLeuAlaSerTrp 1042  
|||||:|||||  
412 .....TGGCCAAAGGAGACAGTGAAGCCCGTG 438  
1043 GluAlaGlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGl 1059  
:|||||:|||||  
439 GAGGTGGAGGAGGAGTCACTGCTGCTTGCCTTGCAACCTCCCCCAAG 488  
1059 yAlaGluGluVal.....LeuLeuHisLeuP 1068  
:|||||:|||||  
489 TGCAGAGCCCTCCGGATCTACTGGATGAACAGCAAGATCTTCACATCA 538  
1068 roPheThrMetValThrGluGlnArgLeuAspAspIleLeuGlyAsn 1084  
:|||||:|||||  
539 AG.....CAGGAGAGCGGTGACGATGGGCCAGAACCGCAAC 576  
1085 LeuSerGlnGlnProGluLeuArgAspLeuTyrSerLysHisLeuVa 1101  
:|||||:|||||  
577 CTCTACTTTGCCAATGTGCTCACTCCGACACCACTCAGACTAC..... 621

Wed Jul 24 11:35:44 2002

```
1101 lAlaGlnLeuAlaGlnGluPheArgSerHisLeuGluHisGlnAspTr 1118
    ||| : : : : : ||| : : : : :
622 .....ATCTGCCAGCCACCTCCAGGACCACAGGA 652
1118 hrLeuLeuLysProSerGluArgThrSerProValThrLeuSerPro 1134
    ||| : : : : : ||| : : : : :
653 CCATCATT.....CAGAAGGAACCCATT..... 675
1135 HisLysHisValSerGlyPheSerSerSerLeuArgThrSerSerThrG 1151
    ||| : : : : : ||| : : : : :
676 .....GACCTCCGGGTCAAGGCCACCC... 696
1151 yAspAlaGlyGlySerArgArgProHisArgLysProThrIleLeuA 1168
    ||| : : : : : ||| : : : : :
697 .....AACAGCATGATGACAGGAGCCGCGCTGCTCT 730
1168 rgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHis 1184
    ||| : : : : : ||| : : : : :
731 TCCCCACCAAC.....TCCAGCAGCCACCTGGTGGCCTTG 765
1185 LeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHi 1201
    ||| : : : : : ||| : : : : :
766 CAGGGGCAG.....CCATTGGTCCCTGGA 788
1201 sCysGluAlaIleGlyHisProArgProThrIleSerTrpAlaArgAsnG 1218
    ||| : : : : : ||| : : : : :
789 GTGATCGCCGAGGCTTCCAGGCCACCATCAATGCTGGCCCCA 838
1218 lyGluGluValGlnPheSerAspArgIleLeuLeuGln...ProAspAsp 1233
    ||| : : : : : ||| : : : : :
839 GTGGCCCCATGCCA...GCTGACCGCTGCACCTACCAAGAACCAACAAG 885
1234 SerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheThrCy 1250
    ||| : : : : : ||| : : : : :
886 ACCCTGACGTGCTGAAGTGGCGGAGGAGTATGCGGAGTACCGCTG 935
1250 sAsnAlaThrAsnAlaGlyTyArgSerValSerIleAlaValThrL 1267
    ||| : : : : : ||| : : : : :
936 CTGGCGGAGAACTCACTGGCAGTGGCCGCGCATCGTACTATGTCACG 985
1267 euAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThr 1283
    ||| : : : : : ||| : : : : :
986 TGGAGGCTGCCCCCTACTGG.....CTG 1008
1284 GluLysProAlaValThrValAspIleGlySerThrIleLysThrValG 1300
    ||| : : : : : ||| : : : : :
1009 CACAAGCCC.....CAGAGCCATCTATATGGGCC 1037
1300 nGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317
    ||| : : : : : ||| : : : : :
1038 AGGAGAGACTCCCGCTGGAGTCCCAAGTCCAGGCGAGGCCCAACCAAG 1087
1317 luValThrTrpPheArgAsnLysSerLysLeuGlySerPro..... 1330
    ||| : : : : : ||| : : : : :
1088 AGGTCACTGGAGAATCAAC.....GGGATCCCTGTGGAGAG 1125
1331 .....HisHisLeuHisGluGlySerLeuLeuLeuTh 1341
    ||| : : : : : ||| : : : : :
1126 CTGGCCAAGACCAGAGTACCGGATTACCGTGGCGCCTGATCCTGAG 1175
1341 rAsnValSerSerAspGlnGlyLeuTyTrSerCysArgAlaAlaAsnL 1358
    ||| : : : : : ||| : : : : :
1176 CAACGTGCGAGCCAGTGCACAAATGTTGACCAATGTGAGGCCGCCAACC 1225
1358 euHisGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspPro 1374
    ||| : : : : : ||| : : : : :
1226 GCGAGGGCTCTGTGGCCAAATGCTACATCTAGCTTGTGTC..... 1266
1375 GlnValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaAlaThrG 1391
    ||| : : : : : ||| : : : : :
1267 CAGCTGCCAGCCCAAG..... 1281
```

```
1391 yProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValL 1408
    ||| : : : : : ||| : : : : :
1282 .....ATCTGACTCGGACAATCAGACGTATACGTGGCTG 1315
1408 euAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisPro 1424
    ||| : : : : : ||| : : : : :
1316 TCAG...GGCAGCAGTGCCTACCTTGTGTGCAAGGCCITTCGGAGCGCCT 1362
    ||| : : : : : ||| : : : : :
1425 ValProAsnIleThrTrpPhe...HisGlyGlyGlnProIleValThrAl 1440
    ||| : : : : : ||| : : : : :
1363 GTGCCAGTGTTCAGTGGCTGGAGGATGGGACAACAGCTGCTTCAGGA 1412
    ||| : : : : : ||| : : : : :
1440 aThrGlyLeuThrHisIleLeuAlaAlaGlyGlnIleLeuGlnValA 1457
    ||| : : : : : ||| : : : : :
1413 CGAACGCTTCTTCCCTAT.....GCCATGGGACCCCTGGGCATTC 1453
1457 laasnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsn 1473
    ||| : : : : : ||| : : : : :
1454 GAGACCTCCAGGCCAATGACACCGGAGCTACTTCTGCTGGCTGCCAAT 1503
    ||| : : : : : ||| : : : : :
1474 GluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTr 1490
    ||| : : : : : ||| : : : : :
1504 GACCAAAACAATGTTACCATCATGCTAACCTGAAGGTTAAAGAT..... 1548
1490 pTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArg 1507
    ||| : : : : : ||| : : : : :
1549 .....GCAACTCAGA 1558
1507 lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsn 1523
    ||| : : : : : ||| : : : : :
1559 TCATCAGGCGCCCGCCAGCACCAATCGAGAGAAGGTTCCAGGGTGACC 1608
    ||| : : : : : ||| : : : : :
1524 ProAlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCy 1540
    ||| : : : : : ||| : : : : :
1609 ...TTCAGTGCAGGCGCTCTTTCACCCCTCTTCAGCCAGCCAGCATCAC 1655
1540 sAsnArgArgAspCys..... 1545
    ||| : : : : : ||| : : : : :
1656 CTGCGTGGGAGGCTCGAGACCTCCAGGAGCTGGGACAGTGCACAAGT 1705
1546 .....ProSerArg.TrpMetValThrSerTrpSerAlaCysTh 1558
    ||| : : : : : ||| : : : : :
1706 ACTTCATAGAGGATGGCGCCTGGTTCATCCACAGCCTGAGTACAGCGAC 1755
    ||| : : : : : ||| : : : : :
1558 rArgSerCysGlyGlyValGlnThrArgValThrCysGlnLysL 1575
    ||| : : : : : ||| : : : : :
1756 CAGG..... 1759
1575 euLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGln 1591
    ||| : : : : : ||| : : : : :
1760 .....GCAACTACAGCTCGTGGCGCAGTACCGAAGTGGTGTGGGAGA 1804
    ||| : : : : : ||| : : : : :
1592 ValAlaLysArgProValAspThrGlnAlaCysAsnGlnLeuCysVa 1608
    ||| : : : : : ||| : : : : :
1805 GTAGGCGACAGC..... 1816
1608 lcluTrpAlaPheSerSerTrpGlyGlnCysasnGlyProCys 1622
    ||| : : : : : ||| : : : : :
1817 .TCTTGTGTGGGAGGCGCTGGCGCGGTGCCAGCGCTGCTGTC 1858
```

seq\_name: /cgn\_5/ptodata/2/ina/5B\_COMB.seq:US-08-427-497E-2

```
seq_documentation_block:
; Sequence 2, Application US/08427497E
; Patent No. 5969124
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
; Patent No. 5969124
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized thereby
; NUMBER OF SEQUENCES: 44
```



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fay, Sharpe, Beall, Fagan,
ADDRESSEE: Minnich & McKee
STREET: 1100 Superior Avenue
CITY: Cleveland
STATE: Ohio
COUNTRY: U.S.A.
ZIP: 44114-2518
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 kb
MEDIUM TYPE: storable
COMPUTER: Compaq Prolinea 5100e
OPERATING SYSTEM: DOS 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 07/904,991
FILING DATE: June 26, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard J.
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: CWR 2 149-3-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 861-5582
TELEFAX: (216) 241-1666
TELEX: (216) 980162
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3774
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: nucleic acids
HYPOTHETICAL: irrelevant
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
INDIVIDUAL ISOLATE: 17-18 week fetus
IMMEDIATE SOURCE:
LIBRARY: Stratagene cDNA Library 936206
CLONE: synthesis of 4 clones
PUBLICATION INFORMATION:
AUTHORS: Hlavin, Mary Louise
AUTHORS: Lemmon, Vance
TITLE: Molecular structure and functional testing of
TITLE: human L1CAM: an interspecies comparison.
JOURNAL: GENOMICS
VOLUME: 11
ISSUE:
PAGES: 416-423
DATE: 1991
RELEVANT RESIDUES IN SEQ ID NO: 1 to 3774
US-08-427-497E-2

alignment_scores:
Quality: 312.00 Length: 815
Ratio: 0.904 Gaps: 30
Percent Similarity: 42.331 Percent Identity: 20.368

alignment_block:
US-10-044-807-2 x US-08-427-497E-2 ..
Align seg 1/1 to: US-08-427-497E-2 from: 1 to: 3774

843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHi 859
|||||
46 CCTGCTGCTTATCCAGATC.....CCCGAGGAATATGA 80

```

```

859 sserProHisIleAlaAlaAlaArgLysValTyrIleGlnThrArgArg 876
::: |||:::
81 AGGACACCATGTGTGGAGCCACCTGTATCAGGAAACAGTCTCCACGGC 130
::: |||:::
876 lnArgLysLeuHisPheValValGlyGlyPheAlaTyrLeuLeuProLys 892
::: |||:::
131 GC.....CTGGTTGTCTTCCCCACA 150
893 ThrAlaValValLeuArgCysProAlaArgValArgLysProLeuIl 909
::: |||:::
151 GATGACATCAGCCTCAAGTGTGAGCCAGTGGCAAGCCCGGAGTGCAGTT 200
::: |||:::
909 eThrTrpGluLysAspGlyGlnHisLeuIleSerSerThrHisValThrV 926
::: |||:::
201 CCCTGGAGCAGGGATGTGTCTTCAAAACCCCAAGGAAGCTGGGTG 250
::: |||:::
926 alAlaProPheGlyTyrLeuLysIleHisArgLeuLysProSerAspAla 942
::: |||:::
251 TGACCGGTGTAC.....CAGTCCGCCCACTCT 276
943 GlyValTyrThrCysSerAlaGlyProAlaArgGluHisPheValIleL 959
||| ::|||
277 GCTTCCTTCACCATCACGGGCAAC.....AACAGCAACTTGTCTCAGAG 320
::: |||
959 sLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeuSerProArgS 976
::: |||
321 GTTCCAGGGCATCTACCGCTGCTTTGCCAGCAATAAGCTGGCACCACCA 370
::: |||
976 erGluGluGluValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeu 992
::: |||
371 TGTCCTCATGATCCGGCTCATGGCGAGGGTGCCTCCCAAG.....411
993 GlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGl 1009
411 .....
1009 uLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAspLeuValS 1026
411 .....
1026 erArgLeuLeuGluGlnGlyTyrProGlyGluLeuLeuAlaSerTrp 1042
||| |||
412 .....TGGCCAAAGAGACAGTGAAGCCCGTG 438
1043 GluAlaGlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGl 1059
|||:::
439 GAGGTGGAGGAAGGGAGTCACTGCTGCTTGCCTTGCACCACTCCCCCAAG 488
1059 yAlaGluGlnVal.....LeuLeuHisLeuP 1068
|||:::
489 TGCAGAGCCTCTCCGGATCTACTGGATGAACAGCAAGATCTTGCACATCA 538
1068 roPheThrMetValThrGluGlnArgArgLeuAspAspIleLeuGlyAsn 1084
::: |||
539 AG.....CAGGACGAGCGGTGACGATGGCGCCAGACGCGCAAC 576
1085 LeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeuVa 1101
|||
577 CTCTACTTTCCTCAATGTCTCACCCTCCGACCAACCACTCAGACTAC.....621
1101 lAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspT 1118
|||:::
622 .....ATCTGCCAGCGCCACTTCCAGGACCAAGCA 652
1118 hrLeuLeuLysProSerGluArgArgThrSerProValThrLeuSerPro 1134
|||:::
653 CCATCAIT.....CAGAAAGGAGCCCAT.....675
1135 HisLysHisValSerGlyPheSerSerSerSerSerSerSerSerThrGl 1151
|||:::
676 .....GACTCCGGGTCAAGGCCACC...696
1151 yAspAlaGlyGlySerArgArgProHisArgLysProThrIleLeuAla 1168

```

```
697      : : : | | | | | | | | : : : | |
1168 rglYsIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHis 1184
      : : : | | | | | | | | : : : | |
731 TCCCCACCAAC.....TCCAGCAGCCACCTGGTGGCCTWG 765
1185 LeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHi 1201
      | | | | | | | | : : : | | | | | :
766 CAGGGGCAG.....CCATTGGTCCTGGA 788
1201 sCysGluAlaIleGlyHisProArgProThrIleSerTrpAlaAsnG 1218
      : | | | | | | | | : | | | | | | | :
789 GTGCATCGCGAGGGCTTCCACGCCACCACATCAATGCTCGGCCCA 838
1218 lYclulGlnValGlnPheSerAspArgIleLeuLeuGln...ProAspAsp 1233
      : : : | | | | | | | | : | | | | | :
839 GTGGCCCCATGCA...GCTGACCGGTGTACCTACCAGCAACCAACAAG 885
1234 SerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheThrCy 1250
      : | | | | | | | | : | | | | | | | :
886 ACCCTGCAGCTGCTGAAGTGGCGAGGAGGATGGCGAGTACCGCTG 935
1250 sAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValThrL 1267
      : | | | | | | | | : | | | | | | | :
936 CTTGGCCGAGAACTCACTGGCGAGTCCCGGCATGGGTACTATCTCACCG 985
1267 euAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThr 1283
      : : : | | | | | | | | : | | | | | :
986 TGGAGGCTGCCCGTACTGG.....CTG 1008
1284 GlutysProAlaValThrValAspIleGlySerThrIleLysThrValGl 1300
      : | | | | | | | | : | | | | | | | :
1009 CACAAGGCC.....CAGAGCCATCTATATGGCC 1037
1300 nGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317
      : | | | | | | | | : | | | | | | | :
1038 AGGAGAGACTGCCCGCTGAGTGCCTCAAGTCCAGGGCAGGCCCAACAG 1087
1317 luValThrTrpPheArgAsnLysSerLysLeuGlySerPro..... 1330
      : | | | | | | | | : | | | | | | | :
1088 AGGTACCTGGAGAAUACA.....GGATCCCTGTGGAGGAG 1125
1331 .....HisHisLeuHisGluGlySerLeuLeuLeuTh 1341
      : | | | | | | | | : | | | | | | | :
1126 CTGGCCAAAGACCAAGATACGGGATTACGGTGGCGCCCTGATCCTGAG 1175
1341 rAsnValSerSerAspGlnGlyLeuTyrSerCysArgAlaAlaAsnL 1358
      : | | | | | | | | : | | | | | | | :
1176 CAACGTGCAGCCCAAGTGCACCAATGGTGACCAATGTGAGGCCGCAACC 1225
1358 euHisGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspPro 1374
      : | | | | | | | | : | | | | | | | :
1226 GGACAGGGCTCTGCTGGCCAAATGCCATACCTACCTGCTGTC..... 1266
1375 GlnValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaAlaThrGl 1391
      : | | | | | | | | : | | | | | | | :
1267 CAGTGCAGCCCAAG..... 1281
1391 yProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValL 1408
      : | | | | | | | | : | | | | | | | :
1282 .....ATPCTGATGGGGAACAATCAGACGATACATGGCTG 1315
1408 euAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisPro 1424
      : | | | | | | | | : | | | | | | | :
1316 TCCAG...GGCAGACTGCTACCTCTCTGCAAGCCCTCGGAGCGCT 1362
1425 ValProAsnIleThrTrpPhe...HisGlyGlnGlnProIleValThrAl 1440
      : | | | | | | | | : | | | | | | | :
1363 GTGCCAGTGTTCAGTGGCTGAGGAGGATGGACACAGAGCTTCAGGA 1412
1440 aThrGlyLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValA 1457
      : : : | | | | | | | | : | | | | | | | :
```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_comb.seq; US-08-506-296B-13

seq\_documentation\_block:  
; Sequence 13, Application US/08506296B  
; Patent No. 6313265  
; GENERAL INFORMATION:  
; APPLICANT: Phillips, Greg  
; APPLICANT: Cunningham, Bruce A.  
; APPLICANT: Crossin, Kathryn L.  
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES  
; NUMBER OF SEQUENCES: 77  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute  
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: U.S.  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/506,296B  
; FILING DATE: 24-JUL-1995

## CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: TSRI 488.0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 554-2937

TELEFAX: (619) 554-6312

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 3888 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 12..3773

US-08-506-296B-13

## alignment\_scores:

Quality: 312.00 Length: 815

Ratio: 0.904 Gaps: 30

Percent Similarity: 42.331 Percent Identity: 20.368

## alignment\_block:

US-10-044-807-2 x US-08-506-296B-13 ..

Align seg 1/1 to: US-08-506-296B-13 from: 1 to: 3888

843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHi 859

|||||

57 CCCTGCTGCTATCCAGATC.....COCGAGGAATATGA 91

859 sSerProHisIleAlaAlaArgLysValTyrIleGlnThrArgArgG 876

|||||

92 AGGACACATGTGTAGGACCACTGTCATCAGGAACTCTCCACGGC 141

876 lnArgLysLeuHisPheValValGlyGlyPheAlaTyrLeuProLys 892

|||||

142 GC.....CTGGTGTCTTCCCCACA 161

893 ThrAlaValValLeuArgCysProAlaArgValArgLysProLeuIl 909

|||||

162 GATGACATCAGCTTCAAGTGTGAGCCAGTGGCAAGCCGAGTGCAGTT 211

909 eThrTrpGluLysAspGlyGlnHisIleLeuSerSerThrHisValThrV 926

|||||

212 CCCTGACGAGGAGTGTCTTCACTTCAACCCAGGAGAGCTGGGTG 261

926 alAlaPropheGlyTyrLeuLysIleHisArgLeuLysProSerAspAla 942

|||||

262 TGACCGGTGAC.....CAGTCCGCCCATCT 287

943 GlyValThrCysSerAlaGlyProAlaArgGluHisPheValIleLy 959

|||||

288 GGTCTCTTCAACATCAGGGCAAC.....AACAGCAACTTGTCTCAGAG 331

959 sLeuIleGlyGlyAsnArgLysLeuValAlaArgProLeuSerProArgS 976

|||||

332 GTTCCAGGGCATCTACCGTGTCTTGGCCACAATAGCTGGGCACGCCA 381

976 erGluGluValLeuAlaGlyArgLysGlyGlyProLysGluAlaLeu 992

|||||

382 TGTCCCATGAGATCCGGCTCATGCCGAGGGTCCCCCAAG. 422

993 GlnThrHisLysHisGlnAsnGlyIlePheSerAsnGlySerLysAlaGl 1009

|||||

422 .....

1009 uLysArgGlyLeuAlaAlaAsnProGlySerArgTyrAspLeuValS 1026

422 .....

1026 erArgLeuLeuGlnGlyTyrProGlyGluLeuLeuAlaSerTrp 1042

423 .....

1043 GluAlaGlnAspSerAlaGluArgAsnThrThrSerGluGluAspProGl 1059

|||||

450 GAGGTGGAGGAGGGAGTCACTGGTTCTGCCTTGCACCCCTCCCAAG 499

1059 yAlaGluGlnVal.....LeuLeuHisLeuP 1068

|||||

500 TGCAGAGCCTCTCCGATCTACTGGATGAACAGCAAGATCTTGCACATCA 549

1068 roPheThrMetValThrGluGlnArgArgLeuAspIleLeuGlyAsn 1084

|||||

550 AG.....CAGGACGAGCGGTGACGATGGCCAGACACGCCAAC 587

1085 LeuSerGlnGlnProGluGluLeuArgAspLeuTyrSerLysHisLeuVa 1101

|||||

588 CTCTACTTTGCCAATGTGCTCCTCCGACAACTCAGACTAC..... 632

1101 lAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisGlnAspT 1118

|||||

633 .....

1118 hrLeuLeuLysProSerGluArgArgThrSerProValThrLeuSerPro 1134

|||||

664 CCATCATT.....CAGAAGGAACCCATT..... 686

1135 HisLysHisValSerGlyPheSerSerLeuArgThrSerSerThrGl 1151

|||||

687 .....

1151 yAspAlaGlyGlyGlySerArgArgProHisArgLysProThrIleLeuA 1168

|||||

708 .....

1168 rgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValValThrHis 1184

|||||

742 TCCCCACCAAC.....TCCAGCAGCCACCTGGTGGCCTTG 776

1185 LeuGlyGlnThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHi 1201

|||||

777 CAGGGGAG.....CCATTGGTCTCTGA 799

1201 sCysGluAlaIleGlyHisProArgProThrIleSerTrpAlaArgAsnG 1218

|||||

800 GTGCATCGCCGAGGCTTCCACGCCCACTCAATGGTGGCCCA 849

1218 lyGluGluValGlnPheSerAspArgIleLeuLeuGln...ProAspAsp 1233

|||||

850 GTGCCCCCATGCCA...GCCAGCGTGTACCTACCAGAACCAACAAG 896

1234 SerLeuGlnIleLeuAlaProValGluAlaAspValGlyPheThrCy 1250

|||||

897 ACCCTGACGTCTGCTGAAAGTGGCGGAGGAGGATGATGGCGAGTACCGCTG 946

1250 sAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerIleAlaValThrL 1267

|||||

947 CCTGGCGGAGAACTCACTGGCAGTGGCCGATGCGTACTATGTACCG 996

1267 euAlaGlyLysProLeuValLysThrSerArgMetThrValIleAsnThr 1283

|||||

997 TGGAGGTGCCCGTACTGG.....CTG 1019

1284 GluLysProAlaValThrValAspIleGlySerThrIleLysThrValGl 1300

|||||

1020 CACAAGCCC.....CAGAGCCATCTATATGGGCC 1048

1300 nGlyValAsnValThrIleAsnCysGlnValAlaGlyValProGluAlaG 1317

wed Jul 24 11:35:44 2002

1049 AGGAGAGACTGGCGCTGGAGTGGCAAGTCCAGGGCAGGCCCAACACAG 1098  
1317 luValThrTrpPheArgAsnLysSerLysLeuGlySerPro..... 1330  
1099 AGGTCACCTGGAGAAATCAAC.....GGGATCCCTGTGGAGGAG 1136  
1331 .....HisHisLeuHisGluGlySerLeuLeuLeuTh 1341  
1137 CTGGCCAAAGACACAGAGTACCGGATTCACGCTGGCGCTGATCTGAG 1186  
1341 rAsnValSerSerAspGlnGlyLeuTyrSerCysArgAlaAlaAsnL 1358  
1187 CAACGTGACGCCAGTCACACATGGTGACCAATGTGAGGCCGCCAACC 1236  
1358 euHisGlyGluLeuThrGlnLeuSerThrGlnLeuLeuLeuLeuAspPro 1374  
1237 GGCAGGGCTCTTCTGGCCCAATGCCATACATCTACGTTGTC..... 1277  
1375 GlnValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaIaThrGl 1391  
1278 CAGCTGCCAGCCAG..... 1292  
1391 yProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValL 1408  
1293 .....ATCCTGACTGGGCAATCAGACGTACATGGCTG 1326  
1408 euAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisPro 1424  
1327 TCCAG...GGCAGCAGTCCCTACCTCTGTCGCAAGGCCCTCGGAGCGCT 1373  
1425 ValProAsnIleThrTrpPhe...HisGlyGlnProIleValThrAl 1440  
1374 GTGCCAGTGTTCAGTGGCTGGAGGAGTGGCAACAGTGTTCAGGA 1423  
1440 aThrGlyLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnVala 1457  
1424 CGAACGCTCTCTCCCTAT.....GCCAATGGGACCTGGGCATTC 1464  
1457 laAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsn 1473  
1465 GAGACCTCCAGGCCCAATGACACCGGACCTACTTCTGCTGGCTGCCAAT 1514  
1474 GluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTr 1490  
1515 GACCAAAACATGTTACCATCATGCTTAACCTGAAGTTAAAGAT..... 1559  
1490 pTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgG 1507  
1560 .....GCAACTCAGA 1569  
1507 lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsn 1523  
1570 TCACCTCAGGGCCCGCCAGCACATCGACAAAGAGTTCCAGGTGACC 1619  
1524 ProAlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCy 1540  
1620 ...TTCAGTGGCAGGCTCTCTTTCACCTTTCACCTCTGTCAGCCCGCAGCATC 1666  
1540 sAsnArgArgAspCys..... 1545  
1667 CTGGGCTGGGAGCGTTCGAGACCTCCAGGAGCTTGGGACAGTGACAAGT 1716  
1546 .....ProSerArg.TripMetValThrSerTyrSerAlaCysTh 1558  
1717 ACTTCATAGAGGATGGGCCCTGGTCATCCACAGCTGACATACAGCGAC 1766  
1558 rArgSerCysGlyGlyGlyValGlnThrArgArgValThrCysGlnLysL 1575  
1767 CAGG..... 1770  
1575 euLysAlaSerGlyIleSerThrProValSerAsnAspMetCysThrGln 1591  
..... 1770

1771 .....GCAACTACAGTGGCTGGCCAGTACCGAACTGGATGTGGTGAGA 1815  
1592 ValAlaLysArgProValAspThrGlnAlaCysAsnGlnGlnLeuCysVa 1608  
1816 GTAGGGCACAGC..... 1827  
1608 lGluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCys 1622  
1828 .TCTTGTGTGGGAGCGCTGGCGCGTGGCCAGCGCTGGTGC 1869  
seq\_name: /cgn2\_6/plodata/2/ina/6B\_COMB.seq:US-09-041-886-24  
seq\_documentation\_block:  
; Sequence 24, Application US/09041886  
; Patent No. 6235872  
; GENERAL INFORMATION:  
; APPLICANT: Bredesen, Dale E.  
; APPLICANT: Rabizadeh, Shariroz  
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/041,886  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 2626  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4608 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..4342  
; US-09-041-886-24  
alignment\_scores:  
Quality: 311.50 Length: 730  
Ratio: 0.986 Gaps: 32  
Percent Similarity: 43.288 Percent Identity: 20.685  
alignment\_block:  
US-10-044-807-2 x US-09-041-886-24 ..  
Align seg 1/1 to: US-09-041-886-24 from: 1 to: 4608  
1148 SerSerThrGlyAspAlaGlyGlySerArgArgProHisArgLysPr 1164  
:::||||| |||||:::|||||:::|||||  
355 GCATCTTAGGAGATCTGGCTCAATATTAGTCGG..... 390  
1164 oThrIleLeuArgLysIleSerAlaAla.....GlnGlnLeuSerA 1178  
|||||:::|||||:::|||||

```

391 .....ACAGCAAAAGTTGCAGTAGCAGGACCACCTGAGGTTCCCTTCAC 433
1178 laSerGluValValThr...HisLeuGlyGlnThrValAlaLeuAlaSer 1193
434 ACAGACAATCTGTACAGCCVTCATGGGAGACACA..... 468
1194 GlyThrLeuSerValLeuLeuHisCysGluAlaIleGlyHisProArgPr 1210
469 .....GTGCTACTCAAGTGTGAAGTCATTGGGAGCCCATGCC 506
1210 oThrIleSerTrpAlaArgAsnGlyGluGlu.....ValGlnPheS 1224
507 AACAAATCCACTGGCAGAGAACCAACAGACCTGACTCCATCCAGGTG 556
1224 erAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaPro 1240
557 ACTCCCGAGTGGTGGTCTGCGCTCTGGAGCATTCGACATCAGCGGACTC 606
1241 ValGluAlaAspValGlyPheTyrThrCysAsnAlaThrAsnAlaLeuG 1257
607 CAACCGGGGACATTTGAATTTACCGATGTCAGTCGAAATCCA..... 651
1257 yTrpAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
652 .....GCCAGCTCAAGACACAGGA..... 669
1274 yThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
670 .....AATGAAGCAGAAGTCAGAACTTTTATCA 696
1291 AspileGly.....SerThrII 1296
697 GATCCAGGACTGCATACACAGCTGTATTTCGCAAGACCATCCAATGT 746
1296 eLysThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyV 1313
747 AGTAGCATTGAAGGAAGATGCTGCTCGGAATGTGTGTCTTGCT 796
1313 alProGluAlaGluValThrTrpPheArgAsnLysSerLysLeu..... 1327
797 ATCCTCCACCAAGTATTACTGTGTTACGGGCGGAGGAAGTCATCCAAC 846
1328 ...GlySerProHisHisLeuHisGluGlySer...LeuLeuLeuThrAs 1342
847 AGGTCTAAAAAGTATCTTTATTGGGTGAAGCAACTTGTCTATCTCCAA 896
1342 nValSerSerAspGlnGlyLeuTyrSerCysArgAlaAlaAsnLeuH 1359
897 TGTGACAGATGATGACAGTGAATGTATACCTGTGTGTGCACATATAAAA 946
1359 isGlyGluLeuThrGluSerThrGlnLeuLeuIleLeuAspProGln 1375
947 ATCAGATAATTAGTCGCTCTGCAGAGCTCACAGTCTGTGGTCCGCCATGG 996
1376 ValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaAlaThrGlyPr 1392
997 TTT.....TT 1001
1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409
1002 AAATCATCTCTCCAAGCTGTATGCGCTATGAAAGCATGGATATTGAGTTTG 1051
1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425
1052 AA.....TGTACAGACTCTCTGGAAGCCTGTG 1077
1426 ProAsnIleThrTrpPheHisGlyGlnProIleValThrAlaThrG 1442
1078 CCCACTGTCAATTGATGAGAAATGGAGATGGTCATTCCTAGT..... 1122
1442 yLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnValAlaAsnL 1459
1123 ....GATTATTTCAGATAGTGGGAGGAAGCAACTACGGATACTTGGG 1168

```

```

1459 euSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAla 1475
1169 TGGTGAAGTCAGATGAAGGCTTTTATCAATGTGTGGCTGAAATAGGCT 1218
1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTrpTrpSe 1492
1219 GGAATGCCAGACCGAGTGCACAGCTCATTTGCCCT..... 1254
1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509
1254 ..... 1254
1509 LnglnProArgLeuArgCysLeuLeuAsnSerThrGluValAsnProAla 1525
1255 .....AAGCCTGCA 1263
1526 HisCysAlaGlyLysValArgProAla.....ValGlnProII 1538
1264 ATCCCAAGCTCCAGTGTCTCTCCCTCGGCTCCAGAGATGTGGTCCCTGT 1313
1538 eAlaCysAsnArgArgAspCysProSerArgTrpMetValThrSerTrps 1555
1314 CTGGTTCCTC.....AGCCGATTGTGCTCGTCTCAGCTGGC 1348
1555 erAlaCysThrArgSerCysGlyGlyValGlnThr..... 1567
1349 GCCCA...CCTGCAGAGAGGAAGGAACATTCAAACTTTCACGCTCTTT 1395
1567 ..... 1567
1396 TTCTCCAGAGAAGGTGACACAGGAGGAGCATTGAATACACACAGGCC 1445
1568 .....ArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerT 1582
1446 TGGTCCCTTCAGCTCAGTGTGGGAAACCTGAGCCAGAACCCATGTACA 1495
1582 hr.....ProValSerAsnAspMet 1588
1496 CCYTTCAGTGTGGCTTACATGAATGGGGACCGGAGAGAGTCTCAA 1545
1589 CysThrGlnValAlaLysArgPro.....Va 1597
1546 CCCATCAAGGTGGCCACACAGCTGAGTTCGCAAGTCCAGGGCCAGTAGA 1595
1597 lAspThrGlnAlaCysAsn.....GlnGlnLeuCysValGlnTrpA 1611
1596 AAACCTGCAAGCTGTATCTACCTACCTACCTCAATTCATTATTACCTGG 1645
1611 laPheSerSerTrpGlyGlnCysAsnGlyProCysIleGlyProHisLeu 1627
1646 AACCCCTCGCTAT.....GCAACGGTCCA..... 1671
1628 AlaValGlnHisArgGlnValPheCysGln..... 1637
1672 ...GTCCAAGTTACAGATTGTCTGCAGCTGAGGTGTCCACAGGAAAGA 1718
1637 ..... 1637
1719 ACAGAAATAGAGGTTGATGGACTATCTTATAAAGTGAAGGCTGAAAA 1768
1637 ..... 1637
1769 AATTCACCAANTATAGTCTTCGATTCAGTTCATTAATCGTAIGTCCG 1818
1638 .....ThrArgAspGlyIleThr.....LeuProSe 1646
1819 GGCGTCTCTACTGATGATATACAGTGGTTACACTTCTGACGTGCCAAG 1868
1646 rGlu.....G 1648
1869 TGCCCCCGCTCAGAAGCTCTCCCTGGAGGTGTCAATTTCAGAGNAGTATCA 1918

```

```
1648 IncysSerAlaLeuProArgProValSerThrGlnAsnCysTrpSerGlu 1664
:: ||| ||||| ||| ::||| ||||| ::|
1919 ASGTAGTGGTGGCTCTCTCCATCAGGAACACAAAATGATTTATACC 1968
:: ||| ||||| ||| ::||| ||||| ::|
1665 AlaCysSerValHisTrpArgVal..... 1672
:: ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1969 GGCTATAAAATTCACACAGAAAGAGCCGCCGAGGGTTCAGATGGAAAC 2018
1673 .....SerLeuTrpThrLeuCysThrAla.....ThrCysG 1683
2019 ACTGGAGCAACAACTCTCTGGTACCTATTCACAGGCTGGGAAAGGAA 2068
1683 lYAsnTrpGlyPheGlnSerArgValGluCysValHisAlaArgThr 1699
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
2069 GTCACTACAGTTCCAG.....GTGCAGCCATGACA 2100
1700 AsnLysAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAs 1716
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
2101 GTCAATGGTACT.....GGACCA...CCTTCCAA 2126
1716 nTrpGlnArgCysAsnIleThrProCysGluAsnMetGlu 1729
2127 CTGTTACTGTCAGAG...ACTCAGAGAATGATCTAGAT 2163
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US94-05277-1
```

```
seq_documentation_block:
; Sequence 1, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zdzienicka, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 BMB UT
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
```

```
; CHROMOSOME/SEGMENT: 18q21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4342
; PCT-US94-05277-1

alignment_scores:
  Quality: 311.50      Length: 730
  Ratio: 0.986        Gaps: 32
  Percent similarity: 43.288      Percent Identity: 20.685

alignment_block:
US-10-044-807-2 x PCT-US94-05277-1 ..
Align seg 1/1 to: PCT-US94-05277-1 from: 1 to: 4608

1148 SerSerThrGlyAspAlaGlyGlySerArgArgProHisArgLysPr 1164
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
355 GCATPCTTTAGGAGATCTGGCTCAATATTAGTCGG..... 390
1164 oThrIleLeuArgLysIleSerAlaAla.....GlnGlnLeuSerA 1178
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
391 .....ACAGCAAAAGTTGCAGTAGCAGGACCACCTGAGGTTCTTTCAC 433
1178 laSerGluValValThr...HisLeuGlyGlnThrValAlaLeuAlaSer 1193
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
434 AGACAGAATCTGTACAGCCCTCATGGGAGACACA..... 468
1194 GlyThrLeuSerValLeuHisCysGluAlaIleGlyHisProArgPr 1210
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
469 .....GTGCTACTCAAGTGTGAAGTCATTTGGGAGCCCATGCC 506
1210 oThrIleSerTrpAlaArgAsnGlyGluGlu.....ValGlnPheS 1224
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
507 AACATCCACCTGGCAGAAAGAACCAAGAGCTGACTCCATCCAGGTG 556
1224 erAspArgIleLeuLeuGlnProAspAspSerLeuGlnIleLeuAlaPro 1240
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
557 ACTCCGAGTGTGGTCTTGCCTCTGGAGCATTCAGATCAGCCGAGTC 606
1241 ValGluAlaAspValGlyPheTrpCysAsnAlaThrAsnAlaLeuGl 1257
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
607 CAACGGGGGACATTCGAATTTACCGATGCTCAGCTCGAATCCA..... 651
1257 yTrAspSerValSerIleAlaValThrLeuAlaGlyLysProLeuValL 1274
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
652 .....GCCAGCTCAAGAACAGGA..... 669
1274 ysThrSerArgMetThrValIleAsnThrGluLysProAlaValThrVal 1290
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
670 .....AATGAGCAGAAAGTCAGAAATTTATCA 696
1291 AspIleGly.....SerThrI 1296
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
697 GATCCAGGACTCCATAGACAGCTGATTTTCGAAAGACCACCAATGT 746
1296 eLysThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyV 1313
: ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
747 AGTAGCCATTGAAGGAAAAGATGCTGCTGGAATGTGTCTGCGCT 796
1313 alProGluAlaGluValThrTrpPheArgAsnLysSerLysLeu..... 1327
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
797 ATCCCCACCAGTTTACTGTGTACGAGGGGAGGAGTCATCCCACTC 846
1328 ...GlySerProHisHisLeuHisGluGlySer...LeuLeuLeuThrAs 1342
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
847 AGGTCTAAAAAGATATCTTTATTGGGTGGAAGCAACTTGTATTATCCAA 896
1342 nValSerSerSerAspGlnGlyLeuTrpSerCysArgAlaAlaAsnLeuH 1359
::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
897 TGTGACAGATGATGACAGTGAATGATATACCTGTGTGTGCACATATAAAA 946
TGTGACAGATGATGACAGTGAATGATATACCTGTGTGTGCACATATAAAA 946
```

1359 isGlyGluLeuThrGluSerThrGlnLeuLeuLeuLeuLeuAspProGln 1375  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
947 ATGAGAAATTATAGTCCCTCGACAGCTCACAGTCTTGTTCCGCCATGG 996  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1376 ValProThrGlnLeuGluAspIleArgAlaLeuLeuAlaAlaThrGlyPr 1392  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
997 TTT..... 1001  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1392 oAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeuValLeuA 1409  
||| ||||| |||::: ::::: |||::: |||::: |||:::  
1002 AAATCATCCTTCACACCTGTATGCCTATGAAGCATGGAATATGAGTTTG 1051  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1409 spProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHisProVal 1425  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1052 AA..... 1077  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1426 ProAsnIleThrTrpPheHisGlyGlnProIleValThrAlaThrGl 1442  
|||::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1078 CCCACTGTGAATTGGATGAAGAATGGAGATGGTCATTCCCTAGT.... 1122  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1442 yLeuThrHisHisIleLeuLeuAlaIaGlyGlnIleLeuGlnValAlaAsnL 1459  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1123 ....GATTATTTCAGATAGTAGTGGGGAAGCAACTTACGGATCTTGGGG 1168  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1459 euSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAla 1475  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1169 TGGTGAAGTCAGATGAAGGCTTTTATCAATGTGGCTGAAAAATGAGGCT 1218  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1476 GlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrrTrpSe 1492  
|||::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1219 GGAATGCCAGACCAGTGCACGCTCAITGTCCCT..... 1254  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1492 rValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArgGlyValG 1509  
1254 ..... 1254  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1509 lnGlnProArgLeuArgCysLeuLeuAsnSerThrGluValAsnProAla 1525  
::: |||||  
1255 .....AAGCCTGCA 1263  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1526 HisCysAlaGlyLysValArgProAla.....ValGlnProIl 1538  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1264 ATCCCAAGCTCCAGTGCCTCCCTTCGGTCCCAGAGATGGTCCCTGT 1313  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1538 eAlaCysAsnArgArgAspCysProSerArgTrpMetValThrSerTrps 1555  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1314 CITGGTTTC.....AGCCGATTGTCCGCTCAGCTGGC 1348  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1555 erAlaCysThrArgSerCysGlyGlyValGlnThr..... 1567  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1349 GCCCA...CCTCAGAAAGCAAAGGAAACAATTCAAACCTTCACGGTCTTT 1395  
1567 ..... 1567  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1396 TTCTCCAGAAAGGTGCACACAGGGACAGCAATTGAATCAACACAGCC 1445  
1568 .....ArgArgValThrCysGlnLysLeuLysAlaSerGlyTleSert 1582  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1446 TGGGTCCCTTTCAGTCACTGTGGGAAAACCTGAAGCCAGAACCCATGTACA 1495  
1582 hr.....ProValSerAsnAspMet 1588  
::: ::::::::::: |||::: ::::: |||::: |||::: |||:::  
1496 CTTTGGAGTTGGCTTACAATGAATGGGGACCGGAGAGAGTCTCAA 1545  
1589 cystThrGlnValAlaLysArgPro.....Va 1597  
::: ||||| |||::: |||::: |||::: |||:::  
1546 CCATCAAGGTGCCACACAGCCTGAGTTGCAAGTTCCAGGGCCAGTAGA 1595  
1597 IaspThrGlnAlaCysAsn.....GlnGlnLeuCysValGlnTrpA 1611  
::: ||||| :::  
1596 AAACCTGCAAGCTGATCTACCTACCTACCTCAATTCATTATTACCTGGG 1645  
1611 lApheSerSerTrpGlyGlnCysAsnGlyProCysIleGlyProHisLeu 1627

1646	AACCCCTGCGCTAT	.....GCAACGGTGCCA	.....	1671
1628	AlaValGlnHisArgGlnValPheCysGln	.....	1637	
1672	...GTCCAAGGTTACAGATGTTCTGCAC	TTGCTGAGGTGTCACAGGAAAGA	1718	
1637	.....	.....	1637	
1719	ACAGAATATAGAGGTGATGGACTATCTTATAAACTGGAAGCCCTCAAA	1768		
1637	.....	.....	1633	
1769	AATTCACCGAATATAGCTTCGTTCGTATTAATCGCTATGTCG	1818		
1638	.....ThrArgAspGlyIleThr	.....LeuProSe	1646	
1819	GGCGTCTCTACTGATGATATAACAGTGGTTACAGTTCTGACGTCCGAC	1868		
1646	rGlu	.....G	1648	
1869	TGCCCGCGCTCAGAACGCTCCCTGGAGTGTTCAATCAAGAAGTATCA	1918		
1648	lnCysSerAlaLeuProArgProValSerThrGlnAsnCysTrpSerGlu	1664		
1919	AAGTTAGCTGGCTGGCTCCCTCCATCAGGAACACAAAATGGATTATTACC	1968		
1665	AlaCysSerValHisTrpArgVal	.....	1672	
1969	GGCTATAAATTCGACACAAAAGACACCGCGGGGTGAGATGGAAC	2018		
1673	.....SerLeuTrpThrLeuCysThrAla	.....ThrCysG	1683	
2019	ACTGGAGCCAAACACTCTGGTACCTATTTCACAGGACTGGAGAAAGAA	2068		
1683	lyAsnTrpGlyPheGlnSerArgValGluCysValHisAlaArgThr	1699		
2069	GTACGATACAGTTTCCAG	.....GTGTGAGCCATGACA	2100	
1700	AsnLysAlaValProGluHisLeuCysSerTrpGlyProArgProAlaAs	1716		
2101	GTCAATGGTACT	.....GGACCA	..CCTTCCAA	2126
1716	nTrpGlnArgCysAsnIleThrProCysGluAsnMetGlu	1729		
2127	CTGGTATATCGAGAG...ACTCCAGAGATGATCTAGAT	2163		

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-506-296B-27

```
seq_documentation_block:
; Sequence 27, Application US/08506296B
```

Patent No. 6313265  
GENERAL INFORMATION.

APPLICANT: Phillips, Greg

APPLICANT: Cunningham, Bruce A.

APPLICANT: Crossin, Kathryn L.

1. TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES  
2. TITLE OF INVENTION: CONTAINING HYDROLYZABLE PEPTIDES

NUMBER OF SEQUENCES: 77

; CORRESPONDENCE ADDRESS:

ADDRESSEE: The Scripps Research Institute

STREET: 10550  
CITY: 12 TOLL

CLIFF: LA JOLLA  
STATE: California

COUNTRY: U

; ZIP: 92037

; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```

;
; SOFTWARE: PatentIn Release #1.0. Version #1 25

```

; CURRENT APPLICATION DATA:

Wed Jul 24 11:35:44 2002

```

FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3943 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 33..3839
US-08-506-296B-27

```

```

alignment_scores:
Quality: 305.00 Length: 955
Ratio: 0.723 Gaps: 41
Percent Similarity: 44.188 Percent Identity: 20.105

```

```

alignment_block:
US-10-044-807-2 x US-08-506-296B-27 ..

```

```

Align seg 1/1 to: US-08-506-296B-27 from: 1 to: 3943

```

```

891 ProLysThrAlaValLeuArgCysProAlaArgValArgLysPr 907
||||| :|||:|||||:|||||:|||||:
195 CCTGAGAGAAATATGTAATACAAATGTAAGCAAAAGGAAACCCCTCC 244
| :|||:|||||:|||||:|||||:
907 oLeuileThrTrpGluLysAspGlyGlnHis.....LeuileSerSert 922
| :|||:|||||:|||||:|||||:
245 TAGCTTCTCTGGACGGCGAATGGAACATCTTTGATATAGATAAAGATG 294
| :|||:|||||:|||||:|||||:
922 hrHisValThrValAlaPro...PheGlyTyrLeuLysIleHisArgLeu 937
:|||||:|||||:|||||:|||||:
295 CACAGGTAAACAATGAACCAAAATTCAGGAACCCCTGTTGTAATATTATG 344
| :|||:|||||:|||||:|||||:
938 LysProSerAspAla.....GlyValTyrThrCysSerAlaG1 950
:|||||:|||||:|||||:|||||:
345 AATGGTGTGAAGGAGAGACATATGAAGGATATACCAAGTGTACAGCAG 394
| :|||:|||||:|||||:|||||:
950 yProAlaArgGluHisPheValIleLysLeuileGlyGlyAsnArgLysL 967
| :|||:|||||:|||||:|||||:
395 GAATCAAGAGAGACGACCAAT.....TCCAACAATA 426
| :|||:|||||:|||||:|||||:
967 euValAlaArgPro...LeuSerProArgSerGluGluValLeuAla 982
:|||||:|||||:|||||:|||||:
427 TTGTTATACGGCCATCAGATCCCGCTTG..... 455
| :|||:|||||:|||||:|||||:
983 GlyArgLysGlyGlyProLysGluAlaLeuGlnThrHisLysHisGlnAs 999
| :|||:|||||:|||||:|||||:
456 .....TGGACTAAAGAAAGAACTAGAACCAAAATCATGTTCCGAGA 493
| :|||:|||||:|||||:|||||:
999 ngGlyIlePheSerAsnGlySerLysAlaGluLysArgGlyLeuAlaAla 1016
| :|||:|||||:|||||:|||||:
494 AGGT..... 497
| :|||:|||||:|||||:|||||:
1016 snProGlySerArgTyrAspAspLeuValSerArgLeuLeuGluGlnGly 1032
| :|||:|||||:|||||:|||||:
498 .....GATTCCCTTAGTACTAAACTGCAGACCTCCTGTT 530
| :|||:|||||:|||||:|||||:
1033 GlyTrpProGlyGluLeuAlaSerTrpGluAlaGlnAspSerAlaG1 1049
| :|||:|||||:|||||:|||||:
531 GGCTTACCACCACTATAATA...TTTGG.....ATGGATAATGCTTT 571

```

```

1049 uArgAsnThrThrSerGluGluAspProGlyAlaGluGlnValLeuLeuH 1066
||||| :|||:|||||:|||||:
572 CCAAGGCGTG.....CCTCAAGTGAAGAGATT..... 599
| :|||:|||||:|||||:|||||:
1066 isLeuProPheThrMetValThrGluGlnArgArgLeuAspAspIleLeu 1082
| :|||:|||||:|||||:|||||:
600 .....TCTCAAGTCTCAATGGAGACCTT 623
| :|||:|||||:|||||:|||||:
1083 GlyAsnLeuSerGlnProGluGluLeuArgAspLeuTyrSerLysHi 1099
| :|||:|||||:|||||:|||||:
624 TATTTTCTAATGTACAAACAGAGACACCGCTGTGGACTAT..... 665
| :|||:|||||:|||||:|||||:
1099 sLeuValAlaGlnLeuAlaGlnGluIlePheArgSerHisLeuGluHisG 1116
| :|||:|||||:|||||:|||||:
666 .....ATCTGCTACGCGAGATTATATCACA 690
| :|||:|||||:|||||:|||||:
1116 lnAspThrLeuLysProSerGluArgArgThrSerProValThrLeu 1132
| :|||:|||||:|||||:|||||:
691 CACAAACTATA.....CAGCAGAAACAACCCATTCT... 722
| :|||:|||||:|||||:|||||:
1133 SerProHisLysHisValSerGlyPheSerSerSerLeuArgThrSerSe 1149
| :|||:|||||:|||||:|||||:
723 .....GTAAAGTCTTTTCAACCAAGCAGTTACA..... 752
| :|||:|||||:|||||:|||||:
1149 rThrGlyAspAlaGlyGlySerArgProHisArgLysProThrI 1166
| :|||:|||||:|||||:|||||:
753 .....GAAAGGCCACCAAGTTC 768
| :|||:|||||:|||||:|||||:
1166 leLeuArgLysIleSerAlaAlaGlnGlnLeuSerAlaSerGluValVal 1182
| :|||:|||||:|||||:|||||:
769 TTCIT..... 773
| :|||:|||||:|||||:|||||:
1183 ThrHisLeuGlyGlnThr.....ValAlaLeuAlaSerGlyThrLe 1196
| :|||:|||||:|||||:|||||:
774 ACACCAATGGCGCAGCAACAAGTAACAAGTGAACCTGAGAGGAATGTCT 823
| :|||:|||||:|||||:|||||:
1196 uSerValLeuLeuHisCysGluAlaIleGlyHisProArgProThrIles 1213
| :|||:|||||:|||||:|||||:
824 T.....TGTGGAATGATCGCAGCAGGATACCCACACCAAGCAATCC 867
| :|||:|||||:|||||:|||||:
1213 erTrpAlaArgAsnGlyGluValGlnPheSerAspArgIleLeuLeu 1229
| :|||:|||||:|||||:|||||:
868 GCTGGATTAAAGAGGGTGGTGAACCTGCCACCAACAGACGGTTTTTTGAA 917
| :|||:|||||:|||||:|||||:
1230 GlnProAspAspSerLeuGlnIleLeuAlaProValGluAlaAspValG1 1246
| :|||:|||||:|||||:|||||:
918 AATTTTAAGAAACTCTCAAGATTATAGACGCTCTCTGAAGCTGACTCTGG 967
| :|||:|||||:|||||:|||||:
1246 yPheTyrThrCysAsnAlaThrAsnAlaLeuGlyTyrAspSerValSerI 1263
| :|||:|||||:|||||:|||||:
968 GAACATACAAATGTACAGCAAGAAATACATGGGTCTTACTCATCATGCA 1017
| :|||:|||||:|||||:|||||:
1263 leAlaValThrLeuAlaGlyLysProLeuValLysThrSerArgMetThr 1279
| :|||:|||||:|||||:|||||:
1018 TTTCGGTAACTGTAAGCTGCCCATACTGGATACACAGCACCAGGAAC 1067
| :|||:|||||:|||||:|||||:
1280 ValIleAsnThrGluLysProAlaValThrValAspIleGlySerThrI1 1296
| :|||:|||||:|||||:|||||:
1068 TTAGTATTGTCT..... 1079
| :|||:|||||:|||||:|||||:
1296 elYstThrValGlnGlyValAsnValThrIleAsnCysGlnValAlaGlyV 1313
| :|||:|||||:|||||:|||||:
1080 .....CCTGGAGAAGATGGGCATTTGATCTCCAGAGCTAATGGCA 1119
| :|||:|||||:|||||:|||||:
1313 alProGluAlaGluValThrTrpPheArgAsnLysSerLysLeuGly... 1328
| :|||:|||||:|||||:|||||:
1120 ACCCAAAACCTAGATAAGCTGTTAAACAAATGGCGTTCCTCATAGCAAT 1169
| :|||:|||||:|||||:|||||:
1329 .....SerProHisHisLeuHisGluGly...SerLeuLeuLeuH 1341
| :|||:|||||:|||||:|||||:
1170 GCCCCAGAAAGATCCTAGCAGAAAGGTAGATGGGGATACCATTTATTTCTC 1219

```



```
1341 rAsnValSerSerAspGlnGlyLeuThrSerCysArgAlaAlaAsnL 1358
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1220 AGCTGTGAAGAACGGTCAAGTGCCTGTTTATCAGTCAATGCTTCTAATG 1269
1358 euHisGlyGluLeuThrGlnSerThrGlnLeuLeuLeuLeu...AspPro 1373
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1270 AGTATGGATACCTGTGCGCAATGATTTGTAATGTTCTTCTGCTGAGCCA 1319
1374 ProGlnValProThrGlnLeuLeuLeuLeuLeuLeuLeuLeuAlaAla 1390
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1320 CCAAGGATCTTAACCTCGCTTAATAAATCTATCAAGTCAATGCA... 1364
1390 rClyProAsnLeuProSerValLeuThrSerProLeuGlyThrGlnLeu 1407
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1365 .....GATAGTCT ..... 1373
1407 alLeuAspProGlyAsnSerAlaLeuLeuGlyCysProIleLysGlyHis 1423
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1374 .....GCATTAATAGACTGTGCTTATTGTTGTTCA 1403
1424 ProValProAsnIleThrTrpPheHisGlyGlyGlnProIleValThrAl 1440
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1404 CCTAAGCCTGAATCGAATGTTTAGGGAGTGAAGAGTAGCATCTTGC 1453
1440 aThrGlyLeuThrHisHisIleLeuAlaAlaGlyGlnIleLeuGlnVal 1457
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1454 AGGA.....AATGAATATGTTTCCATGATATGGAACCTTGGAAATTC 1497
1457 laAsnLeuSerGlyGlySerGlnGlyGluPheSerCysLeuAlaGlnAsn 1473
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1498 CAGTGGCTCAGAAAGATAGTACTGCGCATACACATGTTGTTGCAAGGAT 1547
1474 GluAlaGlyValLeuMetGlnLysAlaSerLeuValIleGlnAspTyrTr 1490
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1548 AATATGGARAGCGCAATGAAGTACAAGTGAAGTTAAGAC..... 1592
1490 pTrpSerValAspArgLeuAlaThrCysSerAlaSerCysGlyAsnArg 1507
1593 .....CCAACGATGA 1602
1507 lyValGlnGlnProArgLeuArgCysLeuLeuAsnSerThrGlnValAsn 1523
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1603 TAATTAACAGCCACATACAAAGTATGATCAGAGATCTGCCAGGCTTCA 1652
1524 ProAlaHisCysAlaGlyLysValArgProAlaValGlnProIleAlaCy 1540
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1653 ...TTGAGTGTGTAATAAACATGATCTTACCTTAATACCAACAGTTAT 1699
1540 s.....AsnArgArgAspCysProSerArgTyrMetVal. 1551
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1700 ATGGCTGAAGACAATAATGAACCTACCAGATGATGAAGGTTTCTAGTTG 1749
1552 .....ThrSerTrpSerAlaCysThrArgSerCysGlyGly 1563
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1750 GTAAGACAACTTTCACATATATGAATGTAATGATAAAGATGATGAACA 1799
1564 GlyValGlnThrArgArgValThrCysGlnLysLeuLysAlaSerGlyII 1580
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1800 TATACTTGATAGTTATTAATCTCTGGACAGTGTTCAGCAAGTGTGT 1849
1580 eSerThrProValSerAsnAspMetCysThrGlnVal.....A 1593
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1850 GCTTACTGTTGTGCTGCTCCCACTCCAGCTATCATTTAGCTCGGC 1899
1593 laLysArgProValAspThrGlnAlaCysAsnGln...GlnLeuCysVal 1608
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1900 CAAATCCACCGCTTGTGGAATTCAGAGTTCAGCTAGAAAGAACATTT 1949
1609 GluTrpAlaPheSerSerTrpGlyGlnCysAsnGlyProCysIle.... 1623
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1950 GAACCTCATGGGTACCAGGAGAAATAACAGTCCATCACTTCACTT 1999
1623 ..... 1623
```

```
2000 TGTGATTGACTATGAAGATGGACTACATGAGCCAGGGGTATGCAATTACC 2049
1624 .....GlyProHisLeuAlaValGlnHisArg..... 1632
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2050 AGACGGAAGTTCCTGGATCTCATACAACATGTACAGTTGAAGTTGCTCCG 2099
1633 .....GlnValPheCysGlnThrArgAspGlyIleth 1643
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2100 TATGTCAACTACTCATCTCGGTGTGATTTGCTCAATGAAATGGTAGAAG 2149
1643 rLeuProSerGluGlnCys.....SerAlaLeuPro. 1653
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2150 TCAGCAAGTGAACCATCTGAACAGTACCTGACAAAGTCCGCAACCCOG 2199
1654 .....ArgProValSerThrGlnAsnCysTrpSerGlu..... 1664
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2200 ATGAATAATCCTTCTAATATGACAAGGATAGGCTCGGAACCTGATAATTG 2249
1664 ..... 1664
2250 GTAATAAGCTGGAGTCTTTAAAGGCTTTCAATGTAATGACCGAGACT 2299
1665 AlaCysSerValHisTrpArgVal.....SerLeuTrpThr. 1676
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2300 CCAATATAAAGTCAGCTGCGGCGCAGAAGATGTTGATGATGATGACGT 2349
1677 ..LeuCysThrAla..ThrCysGlyAsn..... 1684
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2350 CCGTTGTAGTTGCAAAACGTTGCTAAATATATTGTTGTTGTTACACCACT 2399
1685 .....TyrGlyPhe..... 1687
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2400 TTTGTTCCCTATGAATAAAGTACAGGCTTTAAATGACCTGGGATATGC 2449
1688 .....GlnSerArgArgValGluCysValHisAlaArgThrAsnLysAla 1702
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2450 ACCAGGCAATCAGAGGTTATTGGACATTCAGGGGAAGACTTGCCAATGG 2499
1703 ValProGluHisLeuCysSerTrpGlyProArgProAlaAsnTrpGlnAr 1719
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2500 TTGCTCCAGGCAATGTCAGGTTTCATTAACAGCACATTTGGCAAG 2549
1719 qCysAsnIleThr 1723
: |||:||||:||||:||||:||||:||||:||||:||||:||||:
2550 GTGCACTGGGACC 2562
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-383-630-1
seq_documentation_block:
; Sequence 1, Application US/09383630A
; Patent No. 6265632
; GENERAL INFORMATION:
; APPLICANT: Avner Yayon et al.
; TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
; FACTOR RECEPTOR ASSOCIATED
; CHONDRODYSPLASIA
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead+ slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted
; to an ASCII file
; CURRENT APPLICATION DATA:
```

```

> APPLICATION NUMBER: US/09/383,630A
> FILING DATE: 26-Aug-1999
> CLASSIFICATION: <Unknown>
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: <Unknown>
> FILING DATE: <Unknown>
> ATTORNEY/AGENT INFORMATION:
> NAME: Friedmam, Mark M.
> REGISTRATION NUMBER: 33,883
> REFERENCE/DOCKET NUMBER: 1402/2
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 972-3-5625553
> TELEFAX: 972-3-5625554
> TELEX: <Unknown>
> INFORMATION FOR SEQ ID NO: 1:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 5993
> TYPE: nucleic acid
> STRANDEDNESS: double
> TOPOLOGY: linear
> SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-383-630-1

alignment_scores:
Quality: 301.50 Length: 791
Ratio: 0.908 Gaps: 37
Percent Similarity: 41.972 Percent Identity: 21.745

alignment_block:
US-10-044-807-2 x US-09-383-630-1 ..

Align seg 1/1 to: US-09-383-630-1 from: 1 to: 5993

1108 IlePheArgSerHisLeuHisGlnAspThrLeuLeuLysProSerCl 1124
:::||||: :: ||||||| ||| :::: ::::
2243 GTGTTGGGGCGTGGCGGAGCACCACCCGCGCGGCTGCTGC 2292
1124 uArgArgThrSerProValThrLeuSerProHisLysHisValSerGlyP 1141
|||||:::|||||::: ||::: |||:::|
2293 GCCCGGGAGCCGCCAAGTTCAGTGCACCTGGCAGCGGGGTGGCGGAG 2342
1141 heSerSerSerLeuArgThrSerSerThrGlyAspAlaGlyGlyGly... 1156
:::|||||:::|||||::: ||::: |||:::|
2343 CAGCTGGCGCGGTGCGATCCACTCCGCGGGGGAGCTCAGTGTGGCGGG 2392
1157 ..... 1157 .....Se 1157
2393 CCGCCCACTGGGACAGAGGAGACCCCTGGAAACGGCGCCGACAGACGGAG 2442
1157 rArg.....ArgPro..... 1160
||| |||||
2443 CCGCGGTGTGAGTTGGGCTCTAGCGCGCGGATCCCGCGGCTGCAGGA 2492
1161 .....HisArgLysProThrIleLeuArg...LysIles 1171
||| |||::: |||||:::
2493 ATTCGATATCAAGCTTGATCCGCGCGCGCTGCCTGAGGACGCCGCGGG 2542
1171 erAlaAlaGlnLeuSerAlaSerGluValValThrHisLeuGlyGln 1187
:: ||| :::::||||| :::: |||
2543 CCCCCCGCCCGCAVGGCGCCCTCCCTCCCTCGCCCTCGCGCTCGCTG 2592
1188 ThrValAlaLeuAlaSerGlyThrLeuSerValLeuLeuHisCysGlu... 1203
:::|||||:::|||||::: ||| |||
2593 GCGGTGGCCATCGTGGCGCGCGCTCCTCGAGTCTCTTGGGACGGAGCA 2642
1203 ..... 1203 .....
2643 GCGCGTCTGTGGGCGGACGGGCGAGAGTCCCGGCCACAGACCCGCGCCAGC 2692
1204 .....AlaIle..... 1205
2693 AGGAGCACTTGCTTCGGCAGCGGGATGCTGTGGAGCTGAGCTGCCCC 2742

```

1461 yGlySerGlnGlyGlupheSerCysLeuAlaGlnAsnGluAlaGlyValL 1479  
::: |||||:::||||||| ||||| ||||| ||||| ||||| |||||  
3549 TCAGACGCCGGAGTAGTACACCTGCTGGCGGCATTTCTATTGGGT 3598  
  
1478 euMetGlnLysAlaSerLeuValIle.....GlnAsp..TyrTrpTr 1491  
::: ||||| |||||::: |||||::: |||||::: |||||  
3599 CTCATCACTCTGCTGGCTGGTGTCGCCAGCCGAGGAGGAGCTGGT 3648  
l ::: ||| |||||::: .....  
1491 pSerValaspArgLeuAlaThrCys..... 1499  
GAGGCTGACGAGCGGGCAGTGTATGCAGGAATCTCAGTACAGGCT 3698  
1500 ..SerAlaSerCysGlyAsnArgGlyValGlnGlnProArgLeuArgCys 1515  
GGGCTTCTCTGTTTCATCTCTGGTGGTGGCGCTGTGACG..... 3739  
1516 LeuLeuAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValAr 1532  
3740 .....TCTCCGCTCGCGCAGCCGCCCCCAAG 3765  
1532 gProAlaValGlnProIleAlaCysAsnArgArgAspCysProSerArgT 1549  
l ||| ||| |||||::: |||||  
3766 AAGGCCTGGCTCCCCACCGTGCACAAGA..... 3796  
1549 rpMetValThrSerTrpSerAlaCysThrArgSerCysGlyGlyVal 1565  
3797 .....TCTCCCCTTCCCG..... 3811  
1566 GlnThrArgArgValThrCysGlnLysLeuLysAlaSerGlyIleSerTh 1582  
3811 ..... 3811  
1582 rProValSerAsnAspMetCysThrGlnValAlaLysArgPro..... 1596  
|||:::||| ||| |||||  
3812 .....TCAAGCGACAGGTGCTCCTGGAGTCCACGGCTCATGAGTC 3854  
1597 .....ValAspThrGlnAlaCysAsnGlnLeuCysVal 1608  
3855 CAACACACCACACTGTGCGCATCGAAGCTGTCTCAGGGAGGCCCA 3904  
1609 GluTriAlaPheSerSerTrpGlyGlnCys.....AsnGlyPr 1621  
::: ||| |||||::: |||||::: |||||  
3905 CGTGGCCAATGCTCCGAGCTCGAGTGCCTGCCAGCCCAAATGGAG 3954  
1621 oCysIleGlyPro.....H 1626  
3955 CTGCTCGGCCCGCTGACCTGGGCAAGCCCTTGGGAGGCTGCTT 4004  
1626 iLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGly... 1641  
::: |||||::: |||||::: |||||  
4005 CGGCCAGGTGTCATGCGGAGGCCATCGGCATTSACAAGACCGGCGC 4054  
1642 .....IleThrLeuPro....Se 1646  
4055 CCAAGCCTGTACCGTAGCGTCAAGATGCTGAAGACGATGCCACATGC 4104  
1646 rGluGlnCys.....SerA 1651  
::: ||| |||||  
4105 AAGGACCTCTCGGACCTGGTGCTGAGATGGAGATGATGAATGATCGG 4154  
1651 lalaLeuProArgProValSerThrGlnAsnCysTrpSerGluAla..... 1665  
::: ||||| |||||::: |||||  
4155 GAACACAAACAATCATCAACC.....TGCTGGGCGCTGCACGACG 4198  
1666 .....CysSerValHisTrpArgValSerLeuThrTrpThrLeuCysTh 1679  
|||:::||| |||||::: |||||  
4199 GCGGCCCTGTACGTGCTGGTAGTACCGGCCCAAGGGTAACCTCGG 4248  
1679 rAlaThrCysGlyAsnTrpGlyPheGlnSerArgValGluCysValH 1696  
::: |||||::: ||| |||||  
4249 GAGTTCTCGGGCGGGCGCCCGGGCTGGACTACTCTCTCGACAC 4298  
1696 iSAla....ArgThrAsnIysAlaValproGluHisLeuCysSerTrpGly 1711

```

4299 CTGCAAGCCCGCAGGAGCAGTCACTTCAAGG.....ACTGGTGT 4342
1712 ProArgProAlaasnTtp 1717
4343 CCTCTGCTACCAGGTGG 4360

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-383-630-2

seq_documentation_block:
; Sequence 2, Application US/09383630A
; Patent No. 6265632
; GENERAL INFORMATION:
; APPLICANT: Ayner Yayon et al.
; TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
; FACTOR RECEPTOR ASSOCIATED
; CHONDRODYSPLASIA
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; SOFTWARE: Word for Windows version 2.0 converted
; to an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/383,630A
; FILING DATE: 26-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 1402/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5993
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-383-630-2

alignment_scores:
Quality: 301.50 Length: 791
Ratio: 0.908 Gaps: 37
Percent Similarity: 41.972 Percent Identity: 21.745

alignment_block:
US-10-044-807-2 x US-09-383-630-2 ..

Align seg 1/1 to: US-09-383-630-2 from: 1 to: 5993

1108 llepheArgSerHisLeuGluHisGlnAspThrLeuLeuLysProSerGl 1124
:::|||| ::
2243 GTGTTGGGGGTGGCGGAGCACCCCCCAACCCCGCCGGCTGTGC 2292
:::|||| ::
1124 uArgThrSerProValThrLeuSerProHisLysHisValSerGlyp 1141
|||||:::|||||:::

```

Wed Jul 24 11:35:44 2002

2293 GCGCGGCGAGCCCACTTCACGTACTGGCAGCGGGGTGC CGGGAG 2342

1141 heSerSerLeuArgThrSerSerThrdyAspAlaglygly... 1156  
:  
2343 CAGCTGGCGCGTGCATCCACTCGCGCGGGGACTCAGTGTTGGCGG 2392

1157 .....Se 1157  
:  
2393 CCGGCCACTGGACAGAGAGACCCTGGAAAAGCGGCGGAGACGGAG 2442  
1157 rArg.....ArgPro..... 1160  
:  
2443 CCGCGCGTGGTGA GTGGCTCTAGCGCGCGGATCCCCCGGCTGCAGGA 2492

1161 .....HisArgLysProThrIleLeuArg..LysIles 1171  
:  
2493 ATTCGATACAAGCTTGCATCCCGCGCGCTGCTTCAGGACGCCGCGG 2542

1171 erAlaAlaGlnGlnLeuSerAlaserGluValValThrHisLeuGlyGln 1187  
:  
2543 CCCCGCCCCCATGCGGCCCTCTCGCGCTCGCGCTCTGGGTG 2592

1188 ThrValAlaLeuAserGlyThrLeuSerValLeuLeuHisCysGlu.. 1203  
:  
2593 GCCTGCCCATCTGTGGCGCGCTCTCGAGTCTTGGCGCAGGAGCA 2642

1203 ..... 1203  
:  
2643 GCGCGTGTGGCGAGCGAGAGTCCCGGGCCAGAGCCCGGCCAGC 2692

1204 .....Alalie..... 1205  
:  
2693 AGGAGCAGTGTCTTCGGCAGCGGGATGCTGTGGAGCTGAGCTGCC 2742

1206 .....GlyHisProArgProThrIleserTrpAlaArgAsnGlyGl 1219  
:  
2743 CCGCCCGGGGTGTCCATCGGCCCACTCTCTGGTCAAGGATGGCAC 2792

1219 uGluValGlnPheSerAspArgIleLeuGlnProAspSerLeu 1236  
:  
2793 AGGCTGTGCCCTCGAGCGTGTCTGGTGGGCC....CAGCGCTGC 2839

1236 InIleLeuAlaProValgluAlaAspValGlyPheTyThrCys..AsnAl 1252  
:  
2840 AGGTGCTGAATGCTCCACGAGACTCCGGGGCTACAGTGC CGCGGAC 2889

1252 aThrAsnAlaleuclytyrAspSerValserIleala.....ValThrL 1267  
:  
2890 CGGCTCAGCAGCGGTACTGTGCCTTCACTGTGCGGTACAGACGC 2939

1267 euAlaGlyLysProLeuVallylThrSerArgMetThrValIleAsnThr 1283  
:  
2940 TCCATCCTCGGAGATGACGAAGACGGGAGG...ACGAGGTGAGGACA 2986

1284 GluLysProAlaValThrValaspIleGlySerThrIleLysThr..... 1298  
:  
2987 CAGGT.....GTGCACACAGGGGCCCTTACTGCAGCAGCGGC 3023

1299 .....ValGlnGlyValasn...Valt 1305  
:  
3024 CGAGCGGATGCACAAGAAGCTCTGCGCGTCCGGCGCGCAACACCGTCC 3073

1305 hrIleAsnCysGlnValAlaGlyValProGluAlaGluValThrTrpPhe 1321  
:  
3074 GCTTCCGCTGCCACCGCTGCAACCCCACTCCCTCCATCTCTCTGCTGT 3123

1322 ArgAsnLysSerLysLeuGlySerProHisHisLeu..... 1333  
:  
3124 AAGANCGCAGGAGTTCGCGGCCGAGCACCGCATTTGGAGGCATCAAGCT 3173

1334 .....HisGluGlySerLeuLeuThrAsnValSerSerSerAsp 1348  
:  
3174 GCGGCATCAGCATGAGCTGTGTCATGAAGACGCTGGTCCCTCGGAC 3223

1348 InGlyLeuTyrSerCysArgAlaAlaAsnLeuHisGlyGluLeuThrGlu 1364  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3224 GCGCAACTACACCTGGCTGGAGACAAGATTGGCAGCATCCGCAG 3273  
  
1365 SerThrGlnLeuLeuLeuAsp...ProProlGlnValProThrGlnLe 1380  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3274 ACGTACACGCTGCAGCTGCTGGAGCGCTGCCCGCAC..... 3309  
  
1380 uGluAspIleArgAlaLeuLeuAlaAlaThrGlyProAsnLeuProSerV 1397  
||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3310 .....CGGCCATCTCGAGCG.....GGCTGCGCGCCA 3340  
  
1397 alLeuThrSerProLeuGlyThrGlnLeuValLeuAspProGlyAsnSer 1413  
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3341 ACCAGACGGCGGTCTGGCAGCGACCTGGAGTTCCAC..... 3378  
  
1414 AlaLeuLeuCysProIleLysGlyHisProValProAsnIleThrTr 1430  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3379 .....TGCAAGGTGTACAGTCACGCACAGCCCCACATCCAGTG 3416  
  
1430 pPheHis.....GlyGlyGlnP 1436  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3417 GCTCAAGCAGCTGGAGGTGAACGGCAGCAGGTGGCGCGGACGCACAC 3466  
  
1436 rolLeuValThrAlaThrGlyLeuThrHisHisIleLeuAlaAlaGly... 1451  
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3467 CCTACGTTACCGTG.....CTCAAGCGCGGGCGCT 3498  
  
1452 .....GlnIleLeuGlnValAlaAsnLeuSerGI 1461  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3499 AACACCACCGCAAGGAGCTAGAGGTCTCTCCTTGCACAAAGTCACCTT 3548  
  
1461 yGlySerGlnGlyGluPheSerCysLeuAlaGlnAsnGluAlaGlyValL 1478  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3549 TGAGGACGCGGGAGTACACTGCCTGGCGGCAATTCTATTTGGGTTTT 3598  
  
1478 euMetGlnLysAlaSerLeuValIle.....GlnAsp\_TyrTrpTr 1491  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3599 CTCATCACTCTGCTGGCTGGTGGTGCTGCCAGCGAGGAGAGCTGGTG 3648  
  
1491 pServAlaSpArgLeuAlaThrCys..... 1499  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3649 GAGGTGCAGAGCGGGCAGTGTGTATGCAGAAATCTCAGCTACAGGT 3698  
  
1500 ..SerAlaSerCysglyAsnArgGlyValGlnGlnProArgLeuArgCys 1515  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3699 GGGCTTCTCTGTTCATCTCGTGGTGGTGGCGCTGTGACGC..... 3739  
  
1516 LeuLeuAsnSerThrGluValAsnProAlaHisCysAlaGlyLysValAr 1532  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3740 .....CTGCGCGCTGCGCAGCGCCGCCAAG 3765  
  
1532 gProAlaValGlnProIleAlaCysAsnArgArgaspCysProSerArgT 1549  
||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3766 AAAGCCCTGGGCTCCCACCGCTGCACAAGA..... 3796  
  
1549 rpMetValThrSerTrpSerAlaCysThrArgSerCysGlyGlyVal 1565  
||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3797 .....TCTCCCGCTTCCCGC..... 3811  
  
1566 GlnThrArgValThrCysGlnLysAlaSerGlyIleSerTh 1582  
  
3811 ..... 3811  
  
1582 rProValSerAsnAspMetCysThrGlnValAlaLysArgPro..... 1596  
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3812 .....TCAAGCAGAGGTGTCCCTGGAGTCCAACGCGTCCATGAGCTC 3854  
  
1597 .....ValAspThrGlnAlaCysAsnGlnLeuCysVal 1608  
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
3855 CAACACCACTACGCTGCATCAAGGCTGTCTCTCAGGGGAGGGGCCCA 3904

```

1609 GluTrpAlaPheSerSerTrpGlyGlnCys.....AsnGlyP 1621
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3905 CGCTGCCAATGTCCTCGAGCTCGAGCTGCTGCCGACCCCAATGGGAG 3954
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1621 cCysIleGlyPro.....H 1626
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3955 CTGTCTGGGCGCGCTGACCTCGGCAAGCCCTTGGGAGGGTGT 4004
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1626 IsLeuAlaValGlnHisArgGlnValPheCysGlnThrArgAspGly... 1641
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4005 CGGCCAGGTGGTCTGCGGGAGGCATCGCATTCACAAGGACGGCGG 4054
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1642 ..... 1646
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4055 CCAAGCCTGTACCTAGCTGCGGTGAAGATGCTGAAAGACGATGCCACTGAC 4104
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1646 rGluGlnCys.....Sera 1651
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4105 AAGGACCTGTGGGACCTGCTGTGAGATGGAGATGATGAAGATGATCGG 4154
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1651 laLeuProArgProValSerThrGlnAsnCysTrpSerGluAla..... 1665
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4155 GAACACAAAAACATCAAC.....TGTGGGCGCTGCAGCAGG 4198
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1666 .....CysSerValHisTrpArgValSerLeuTrpThrLeuCysTh 1679
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4199 GCGGGCCCTGTACGTGCTGTGAGTACGCGGCCAAGGGTAACCTGCGG 4248
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1679 rAlaThrCysGlyAsnTyrGlyPheGlnSerArgArgValGluCysValH 1696
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4249 GAGTTTCTGGGGCGCGCGCGCGCGCGCTGGACTACTCTTCGACAC 4298
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1696 isAla....ArgThrAsnLysAlaValProGluHisLeuCysSerTrpGly 1711
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4299 CTGCAAGCCCGCGAGGAGGAGCTCACTCAAGG.....ACCTGGTGT 4342
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1712 ProArgProAlaAsnTrp 1717
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4343 CCTGTGCTACCAAGTGG 4360

```

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-427-497E-3

```

seq_documentation_block:
; Sequence 3, Application US/08427497E
; Patent No. 5969124
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
; Patent No. 5969124
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; MEDIUM TYPE: storable
; COMPUTER: Compaq Prolinea 5100e
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08427,497E
; FILING DATE: April 24, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 07/904,991
; FILING DATE: June 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mindich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: CWR 2 149-3-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3189
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acids
; HYPOTHEICAL: irrelevant
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: 17-18 week fetus
; IMMEDIATE SOURCE:
; LIBRARY: Stratagene cDNA Library 936206
; CLONE: 3.1
; PUBLICATION INFORMATION:
; AUTHORS: Blavin, Mary Louise
; AUTHORS: Lemmon, Vance
; TITLE: Molecular structure and functional testing of
; TITLE: human L1CAM: an interspecies comparison.
; JOURNAL: GENOMICS
; VOLUME: 11
; ISSUE:
; PAGES: 416-423
; DATE: 1991
; RELEVANT RESIDUES IN SEQ ID NO: 548 to 3736
; US-08-427-497E-3

```

alignment\_scores:

```

Quality: 287.00 Length: 522
Ratio: 1.206 Gaps: 21
Percent Similarity: 45.594 Percent Identity: 23.755

```

alignment\_block:

US-10-044-807-2 x US-08-427-497E-3 ..

Align seg 1/1 to: US-08-427-497E-3 from: 1 to: 3189

```

1135 HisLysHisValSerGlyPheSerSer.....Se 1144
      ||| ||| ||| :||| :||| :||| :||| :||| :|||
81 CAGCCCACTTCCAGGACCAAGGACCATTCAGAAAGAACCATTTGA 130
1144 rLeuArgThrSerSerThrGlyAspAlaGlyGlySerArgArgProH 1161
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
131 CCTCCGGGTCAAGGCCACC.....AACAGCATGATTG 162
1161 isArgLysProThrIleLeuArgLysIleSerAlaGlnGlnLeuSer 1177
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
163 ACAGGAAGCGCGCGCTGCTTCCCAACCAAC.....TCC 197
1178 AlaSerGluValValThrHisLeuGlyGlnThrValAlaLeuAlaSerGI 1194
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 AGCAGCCACTGTGTGCTTGCCTTGCAGGGGCGAG..... 227
1194 yThrLeuSerValLeuHisCysGluAlaIleGlyHisProArgProT 1211
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
228 .....CCATTGGTCTGTGAGTGCATCGCCGAGGGCTTCCACGCCCA 270
1211 hrIleSerTrpAlaArgAsnGlyGluGluValGlnPheSerAspArgIle 1227
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
271 CCATCAATGGTGGCCCGCCAGTGGCCCATGCCA...GCTGACCGGTGC 317

```

seq name: /cqn2 6/ptodata/2/ina/5A\_COMB.seq:US-07-862-021B-11

: FEATURE:

```

1645 .....::AC1646
530 atpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV547
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1647 ATGGTCCCCCGTGTATCTTCCTGTGAATGGGAACCGCATCTAGA...1692
547 aAlArgCysGlnValLeuSerPheSerGlnSerValAlaAspLeuPro563
      .....|||
1693 .....GAGAGATGTATGAAGCAATTCCTCC1716

```

564 IleAsp.....GluCysGluGlyProLysProAlaSerGlnArgAlaCys 578  
||| ||||| |||  
1717 GAAGATGGCTCTATGTGCTAAAGTGCCTACTCAAGAAACTCAGAAATGTAT 1766  
578 sTyralaClyProCysSerGlyGlu.....IleProGluPheAsn 592  
||||| ||||| |||||  
1767 TGTAAATGAGAAATGCTCCCTACAGCAGCTCCTGTGTACCGAATGGGA. 1815  
592 roAspGluThrAspGlyLeuPheClyGlyLeuGlnAspPheAspGluLeu 608  
1815 ..... 1815  
609 TyrAspTrpGluTrpGluGlyPheThrLysCysSerGluSerCysGlyG1 625  
||||| ||||| |||||  
1816 ..CAGTGGAT.....GAATGCAGTGCCTAGCTGTGGCAC 184  
625 yGlyValGlnGluAlaValSerCysLeuAsnLysGlnThrArgGluP 642  
||||| ||||| |||||  
1848 AGGAATGAAAGCGA.....CACAGAATGATCAAGATGACT.....C 1885  
642 roAlaGluGluAsnLeuCys.....ValThrSerArgArgProGln 656  
||||| ||||| |||||  
1886 CTGCTGATGATCTATGTCAAGGCAGAAACTACAGCGCAGAGAAATGC 1935  
657 LeuLeuLysSerCysAsnLeuAspProCys...ProAlaArgTrpGluI 672  
||||| ||||| |||||  
1936 ATGATGCCGAATGCCATACTATCCCTCCTCTATCCCATGG..... 1980  
672 eGlyLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnTh 689  
1981 .TCTGAATGGAGCGACTGCAGCGTGACATGTGGGAGGAATGCGAACC 2029  
689 rGAspValPheCysSerHisLeuLeuSerArgGluMetAsnGluThrVal 705  
||| ||||| |||||  
2030 GG.....CAAAGATGCTGAAATCTGCAGCT 2055  
706 IleLeuAlaAspGluLeuCysArgGlnProLysProSerThrValGlnAl 722  
||||| ||||| |||||  
2056 GAGCTTGGAGAC.....TGCATGAG...GAATGGAGCAGCAGAGAA 2096  
722 aCysAsnArgPheAsnCysPro.....ProAlaTrpTrp 734  
||| ||||| |||||  
2097 ATGCTGCTACCTGAATGCCCATTTGCTGTGAGTAAAGGATGG.... 2142  
734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750  
:||||| ||||| |||||  
2143 ..TCCAGTGGTCCAGTGCATACCTCCTGTGGGAGGGCCACATGATC 2190  
751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGlu 767  
||| ||||| |||||  
191 AGAACAGAATGAUCAAATAGAACCCAGATTTGGAGA...ACAGCATG 2237  
767 uProGluThrPhe.....Cys..... 772  
||||| |||||  
1238 CCCAGAAATCTGCCAACGCTACTAAATGTGAGTAGTAAAGGAATGCTGAGAG 2287  
773 .....SerAlaSerLysProAla 778  
:||| |||||

Wed Jul 24 11:35:44 2002

```
2288 GCCAGGTATGGAAGAGCGCTTGAAGAGCGCCGCGGAGAAAGAAGA 2337
779 CysGlnGlnAlaCysLysLysAsp.....CysPr 789
:::||||| |||::: |||
2338 AGTGAACAAGCAAAAAAATATGTAATAGCAATATCCAGTTGTAG 2387
789 o....SerGluTrpLeuLeuSerAspTrpThrGluCysSerThrSerCysG 805
::: ||| |||::: ||| ||| ||| ||| ||| ||| ||| |||
2388 GCTGAACACATGG.....ACTGCTGGACAGAAATGTTCTACACTCTGTG 2431
805 lyGluGlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThr 821
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2432 CAGGTGCAATTCAGGCGCTACATGATGCTAAAGAAGAGCTCCAAAAGC 2481
822 GlyLeuSerThrValValAsnSerThrLeuCysProProLeuProPheSe 838
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2482 .....ACTCAGTTTACTAGCTGCAAGAC.....AA 2507
838 rSerSerIleArgProCysMetLeuAlaThrCys 849
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2508 AAGAGCTAAGACAGTACGTTCATCTTGT 2541
```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-313-288B-11

```
seq_documentation_block:
; Sequence 11, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
; US-08-313-288B-11
```

alignment\_scores:
Quality: 277.50 Length: 628
Ratio: 1.110 Gaps: 39
Percent Similarity: 39.809 Percent Identity: 21.975

```
alignment_block:
US-10-044-807-2 x US-08-313-288B-11
Align seg 1/1 to: US-08-313-288B-11 from: 1 to: 3226

278 LysIleArgAsnSerGlySerAlaAspSerThrValGlnPheIlePheTy 294
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1243 AAGATTAGACCACTTACAAGCTTAGATCACCT...CAGAGTCCATTTTA 1289
294 rGlnProIleIleHisArgTrpArgGluThrAspPhePheProCysSerA 311
||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1290 TGATCCAGAA..... 1299
311 laThrCysGlyGlyGlyTyTrpGlnLeuThrSerAlaGluCysTyAspLeu 327
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1300 .....GGAGGATCTATCAAGCTTGATGCC..... 1323
328 ArgSerAsnArgValValAlaAsp.....GlnTyTrpCy 338
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1324 .....AGAGTCGTGCTTGAAGAATTCACGCAAGGGGGAGCAGTG 1364
338 sHisTyTrpProGluAsnIle.....LysProL 348
||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1365 CAACITCGTACTGATAACATAGATGATATGTCGACAGCTAGCACCAAG 1414
348 ysProLysLeuGlnGluCysAsnLeuAspProCysProAlaSerAspGly 364
||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1415 AAGAAAAAGAGAGATGATACCCCTGAGACCTGCATATATCAAAAC... 1461
365 TyrLysGlnIleMetProTyAspLeuTyHisProLeuProArgTrpG1 381
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1462 .....TGG... 1464
381 uAlaThrProTrpThrAlaCysSerSer...CysGlyGlyGlyLeG 397
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1465 .....TCCCCCTGTCAGCTGCAGCTCCTCTACTCTGAGAAGGCAAGA 1510
397 InSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413
:: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1511 GGATGAGG..... 1518
414 SerValGluGluTrpLysCysMetTyTrpProLysMetProIleAlaG1 430
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1519 .....CAGAGAATGCTTAAAGCTCAGCTGAGCTCAGTGT 1553
430 nProCysAsnIlePheAspCysProLysTrpLeuAlaGlnGluTrpSerP 447
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1554 GCCC.....TGCTCTGAT.....ACCCAAGATTTCAGC 1582
447 roCysThrValThrCysGlyGlnGlyLeuArgTyArgValValLeuCys 463
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1583 CATGC.....ATGGGTCCAGC.....TGC 1602
464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1603 AGTGATGAAGATGTT..... 1617
480 oHisIleLysGluGluCysIleValProThrProCysTyTrpLysProLysG 497
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1618 .....TCAACTTGCATGATGTTCT..... 1635
497 luLysLeuProValGluAlaLysLeuProTrpPheLysGlnAlaGlnGlu 513
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1636 .....GACTGGATT..... 1644
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530
:: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1645 .....AC 1646
530 atrpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1647 ATGGTCCCTCTAGTCTTCTGCTGGAATGGGAACGCATCTAGA.... 1692
```



```

547 aArgCysGlnValLeuLeuSerPheSerGlnSerValAlaAspLeuPro 563
1693 .....:|||||:|
564 IleAsp.....GluCysGluGlyProLysProAlaSerGlnArgAlaCys 578
1717 GAAGATGGCTCTATGTGCAAGAGCTCTATGAGAAACCTGAGAAATGTAT 1766
578 stfAlaGlyProCysSerGlyGlu.....IleProGluPheAsn 592
1767 TGTAAATGAGGAATGCTCCCTGACAGCTGCTGTGTACCGAATGGGA. 1815
592 roAspGluThrAspGlyLeuPheGlyGlyLeuGluAspPheAspGluLeu 608
1815 .....
609 TyrAspTrpGluTyrGluGlyPheThrLysCysSerGluSerCysGlyG 625
1816 ...GAGTGGAT.....GAATGAGTGTAGTGTGGCAGC 1847
625 yGlyValGlnGluAlaValSerCysLeuAsnLysGlnThrArgGluP 642
1848 AGGAATGAAAGGGGA.....CACAGAATGATCAAGATGACT.....C 1885
642 roAlaGluGluAsnLeuCys.....ValThrSerArgArgProGln 656
1886 CTGCTGATGGATCTATGTGCAAGCGCAGAACTACAGAGCGAGAAATGC 1935
657 LeuLeuLysSerCysAsnLeuAspProCys...ProAlaArgTrpGlu 672
1936 ATGATGCCGAATGCCATCTATTCCTGCTTCTATCCCATGG..... 1980
672 eGlyLysTrpSerProCysSerLeuThrCysGlyValGlyLeuGlnTh 689
1981 ..TCGAATGGAGGAGCTGACGCTGACATGTGGGAAGGAATGCAACCC 2029
689 rgAspValPheCysSerHisLeuLeuSerArgGluMetAsnGluThr 705
2030 GG.....:|||||:|
706 IleLeuAlaAspGluLeuCysArgGlnProLysProSerThrValGln 722
2056 GAGCTTGGAGAC.....TGCAATGAG...GACTGGAGCAGCAGAGAA 2096
722 aCysAsnArgPheAsnCysPro.....ProAlaTrpTyrP 734
2097 ATGATGCTACCTGAATGCCCAATTGACTGTGAGCTAACGGAGTGG.... 2142
734 roAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValGlnLys 750
2143 ..TCCCAAGTGGTCCGAGTGCATACCTCTCTGTGGGAAGGCCACATG 2190
751 ArgGluValLeuCysLysGlnArgMetAlaAspGlySerPheLeuGlu 767
2191 AGAACAAAGATGATCAAAATAGAACACACAGTGTGGAGGA...ACAG 2237
767 uProGluThrPhe.....Cys..... 772
2238 CCCAGAAACGTCCCAACGTACTAAATGTCGAGTAAGGAATGCTGAGAG 2287
773 .....SerAlaSerLysProAla 778
2288 GCCCAGGTATGGAAGACGCGTGTGAAGAGCCCGGGAGAAAGAA 2337
779 CysGlnGluAlaCysLysLysAspAsp.....CysPr 789
2338 AGTGAACAAGCAAAAAAATATGTATGACAAATATCCAGTTTGTAG 2387
789 o....SerGluTrpLeuLeuSerAspTrpThrGluCysSerThrSer 805
2388 GCTGAACCACTGG.....ACTGCTTGGACAGAAATGTTCTACACTCTG 2431

```

```

805 lyGluGlyThrGlnThrArgSerAlaIleCysArgLysMetLeuLysThr 821
2432 GAGGTGAATTCAGGACCGCTACATGATGTTAAAGAAGAGTCCAAAAGC 2481
822 GlyLeuSerThrValValAsnSerThrLeuCysProProLeuProPheSe 838
2482 .....:|||||:|
838 rSerSerIleArgProCysMetLeuAlaThrCys 849
2508 AAAGGAGCTAAGAGCAATGTAACGTTTCATCCTTGT 2541
seq_name: /cgn2_6/ptodata/2/ina/PCtus_COMB.seq: PCT-US93-03164-11
seq_documentation_block:
; Sequence 11, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
; PCT-US93-03164-11

```

```

alignment_scores:
Quality: 277.50      Length: 628
Ratio: 1.110        Gaps: 39
Percent Similarity: 39.809  Percent Identity: 21.975

```

```
alignment_block:
US-10-044-807-2 x PCT-US93-03164-11

```

```
Align seg 1/1 to: PCT-US93-03164-11 from: 1 to: 3226

```

```
278 LysIleArgAsnSerGlySerAlaAspSerThrValGlnPheIlePheTy 294

```

1243 AAGATTAGACCACTTACAGCTTATAGTACACCT...CAGAGTCCATTTTA 1289  
294 rGlnProIlelleHisArgTrrArgGluThrAspPheProCysSerA 311  
1290 TGATCCAGAA... 1299  
311 laThrCysGlyGlyGlyTyrGlnLeuThrSerAlaGluCysTyrAspLeu 327  
1300 .....GGAGCATCTATCAAGCTTGTAGCC... 1323  
328 ArgSerAsnArgValValAlaAsp.....GlnTyrCys 338  
1324 .....AGAGTCGTCTGTAAGAATTCACCAAGGGGGAGCAGTG 1364  
338 sHisTyrTyrProGluAsnIle.....LysProL 348  
1365 CACATTCGTACCTGATACATAGATGATATTGTGCAGACACCTAGCACCAAG 1414  
348 ysProLysLeuGlnGluCysAsnLeuAspProCysProAlaSerAspGly 364  
1415 AAGAAAAAGAAAGATGATACCCCTGCAGACCTGATATATCAAC... 1461  
365 TyrLysGlnIleMetProTyrAspLeuTyrHisProLeuProArgTrrpG1 381  
1462 .....TGG... 1464  
381 uAlaThrProTrrThrAlaCysSerSer...CysGlyGlyGlyIleG 397  
1465 .....TCCCCCTGGTCAGCTGCAGCTCCCTACCTGTGAGAAGGCAAGA 1510  
397 lnSerArgAlaValSerCysValGluGluAspIleGlnGlyHisValThr 413  
1511 GGATGAGG... 1518  
414 SerValGluGluThrLysCysMetTyrThrProLysMetProIleAlaG1 430  
1519 .....CAGAGATGCTTAAGCTCAGCTGCACCTCAGCTG 1553  
430 nProCysAsnIlePheAspCysProLysTrrpLeuAlaGlnGluTrrpSerP 447  
1554 GCCC.....TGTCCTGAT.....ACCCAAGATTTTCAGC 1582  
447 roCysThrValThrCysGlyGlnGlyLeuArgTyrArgValValLeuCys 463  
1583 CATCC.....ATGGGTCCAGGC.....TGC 1602  
464 IleAspHisArgGlyMetHisThrGlyGlyCysSerProLysThrLysPr 480  
1603 AGTGATGAAGATGT..... 1617  
480 oHisIleLysGluGluCysIleValProThrProCysTyrLysProLysG 497  
1618 .....TCAACTTGATGATGCT..... 1635  
497 luLysLeuProValGluAlaLysLeuProThrPheLysGlnAlaGlnGlu 513  
1636 .....GACTGGATTT..... 1644  
514 LeuGluGluGlyAlaAlaValSerGluGluProSerPheIleProGluAl 530  
1645 .....AC 1646  
530 atrpSerAlaCysThrValThrCysGlyValGlyThrGlnValArgIleV 547  
1647 ATGGTCCCCCTGTAGTGTCTCTGGAATGGGAACCGCATCTAGA... 1692  
547 alArgCysGlnValLeuSerPheSerGlnSerValAlaAspLeuPro 563  
1693 .....GAGATATGTAAAGCAATTCGCC 1716  
564 IleAsp.....GluCysGluGlyProLysProAlaSerGlnArgAlaCy 578  
1717 GAAGATGGCTCTATGTGCAAAAGTGCCTACTGAAGAACTGAGAAATGTAT 1766

838 rSerSerIleArgProCysMetLeuAlaThrCys 849  
:::~::~:~::|||:::  
2508 AAAGGAGCTAAGAGCATGTACGTTTCATCCTTGT 2541



OM of: US-10-044-807-2 to: EST:\* out\_format : pfs

Date: Jul 24, 2002 5:14 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q/cgn2.1/USPTO.spool/US10044807/runat\_22072002.153741.13644/app\_query.fasta\_1.1847  
-DB=EST -QWMT=Fastap -SUFFIX=rst -CAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.500 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DEEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -MODALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MTN=0  
-ALIGN=15 -MODEL=LOCAL -OUTEXT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US10044807 @CGL1.1.4168  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPYX -WAIT -THREADS=1

Search information block:

Query: US-10-044-807-2

Query length: 1762

Database: EST\*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 2791.330000

score\_list:

Sequence	Strd Orig	zScore	EScore Len	Documentation
gb_est1:BM476141	+ 1614.00 2227.34	8.1e-115	1009	! BM476141 AGNCOURT_6478885 NIH
gb_est2:BE888902	+ 1175.50 1618.86	6.3e-81	850	! BE888902 601513828F1 NIH_MGC_71
gb_est1:BB633755	+ 1102.00 1520.30	2.0e-75	622	! BB633755 RIKEN full-le
gb_est1:AI917724	- 1044.00 1440.64	5.4e-71	568	! AI917724 tllc08.xl NCI_CGAP_G
gb_hic:AK020115	+ 1037.00 1424.31	4.3e-70	979	! AK020115 Mus musculus 12 days e
gb_est1:BB659292	+ 998.00 1374.77	2.5e-67	660	! BB659292 BB659292 RIKEN full-le
gb_est2:BI523618	+ 958.00 1320.62	2.6e-64	581	! BI523618 603051757F1 NIH_MGC_12
gb_est1:AI342006	- 956.00 1319.13	3.1e-64	522	! AI342006 q98d10.xl NCI_CGAP_G
gb_est1:BB643318	+ 942.00 1296.09	6.0e-63	700	! BB643318 BB643318 RIKEN full-le
gb_est1:BB478282	+ 934.00 1286.44	2.1e-62	619	! BB478282 BB478282 RIKEN full-le
gb_est1:BB620571	+ 932.50 1284.08	2.8e-62	633	! BB620571 BB620571 RIKEN full-le
gb_est1:BB612331	+ 888.00 1222.04	8.0e-59	637	! BB612331 BB612331 RIKEN full-le
gb_est1:AI261611	+ 869.00 1198.30	1.7e-57	509	! AI261611 qz35h01.xl NCI_CGAP_K
gb_est1:BB193444	+ 842.00 1157.88	3.0e-55	643	! BB193444 BB193444 RIKEN full-le
gb_est1:BE628435	+ 839.00 1155.59	4.0e-55	550	! BE628435 uu09f09.Y2 Soares mam
gb_est1:BF058634	- 826.00 1139.94	3.0e-54	449	! BF058634 7k24d09.xl NCI_CGAP_O
gb_est1:BB615232	+ 811.00 1116.50	6.1e-53	555	! BB615232 BB615232 RIKEN full-le
gb_hic:AK020361	+ 775.00 1061.28	7.2e-50	846	! AK020361 Mus musculus adult mal
gb_est2:BF510853	- 765.50 1052.49	2.2e-49	586	! BF510853 UI-H-B14-4pm-e-10-0-U
gb_est2:BG750735	- 759.00 1043.03	7.5e-49	606	! BG750735 602706924F1 NIH_MGC_4
gb_est2:N33407	- 750.00 1033.99	2.4e-48	454	! N33407 yv41f11.s1 Soares meland
gb_est2:BF11214	- 713.00 978.25	3.0e-45	644	! BF11214 7044e07.xl NCI_CGAP_Lu
gb_est1:AI825646	- 709.00 979.06	2.7e-45	380	! AI825646 wbf5b10.xl NCI_CGAP_G
gb_est1:AA482392	+ 693.50 956.18	5.2e-44	423	! AA482392 zv05h07.Y1 Soares huma
gb_est2:BE631939	+ 693.00 941.26	9.7e-44	600	! BE631939 uu09f09.xl Soares mam
gb_est1:AI750970	- 683.00 937.12	3.3e-43	419	! AI750970 cn06h02.Y1 Normal Huma
gb_est2:BF445136	- 651.00 892.17	1.9e-40	631	! BF445136 nad21a04.xl NCI_CGAP_L
gb_est1:AW120520	- 607.00 832.81	3.8e-37	539	! AW120520 UI-M-BH1-ana-b-01-0-U
gb_est1:AI459225	- 602.50 824.70	1.1e-36	628	! AI459225 tkl1b03.xl NCI_CGAP_Lu
gb_est1:BB839998	- 593.00 801.79	2.1e-35	442	! BB839998 BB839998 RIKEN full-le
gb_est1:AW120794	- 573.00 787.60	1.3e-34	452	! AW120794 UI-M-BH2.3-any-h-08-0
gb_est2:BF67952	+ 555.00 763.28	2.9e-33	425	! BF67952 CM1-CN0061-201200-673-
gb_est2:W47220	+ 548.00 757.23	6.2e-33	313	! W47220 zc39a07.Y1 Soares_sense
gb_est1:AA284278	+ 545.00 725.64	3.6e-31	302	! AA284278 zc39a07.Y7 Soares_sense
gb_est1:AA971699	- 461.00 629.64	8.0e-26	534	! AA971699 op95a12.s1 NCI_CGAP_Lu
gb_est2:W47029	- 457.00 626.49	1.2e-25	437	! W47029 zc39a07.Y1 Soares_sense
gb_gss:CN503V8J	+ 430.50 579.80	4.8e-23	983	! AI262108 Tetraodon nigroviridis
gb_est2:W44869	+ 427.00 592.32	9.6e-24	233	! W44869 yv40f11.xl Soares meland
gb_hic:BC003269	+ 426.50 561.03	5.3e-22	2931	! BC003269 Mus musculus, Similar

gb\_est1:AW027573 - 420.50 571.00 1.5e-22 643 ! AW027573 wv74h06.xl Soares t  
gb\_est2:BI821835 + 418.50 565.81 2.9e-22 785 ! BI821835 603035722F1 NIH\_MGC  
gb\_est1:AL665815 + 409.50 549.89 2.2e-21 1039 ! AL665815 AL665815 direction  
gb\_est2:BI733795 + 408.00 553.20 1.4e-21 746 ! BI733795 603351940F1 NIH\_MGC  
gb\_est2:BI852900 + 407.00 548.68 2.6e-21 861 ! BI852900 603379705F1 NIH\_GCA  
  
seq\_name: gb\_est2:BM476141  
  
seq documentation block:  
LOCUS BM476141 1009 bp mRNA linear EST 05-FEB-2002  
DEFINITION AGNCOURT\_6478885 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:5558669  
5', mRNA sequence.  
ACCESSION BM476141  
VERSION BM476141.1 GI:18525183  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1009)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabps@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM12282 row: a column: 06  
High quality sequence stop: 647.  
  
FEATURES  
Location/Qualifiers  
1..1009  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5558669"  
/clone\_lib="NIH\_MGC\_88"  
/tissue\_type="duodenal adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: small intestine; Vector: pCMV-SPORT6;  
Site.1: NotI; Site.2: SalI; Cloned unidirectionally;  
oligo-dt primed. Average insert size 1.767 kb. Library  
enriched for full-length clones and constructed by Life  
Technologies. Note: this is a NIH\_MGC Library."  
  
BASE COUNT 234 a 299 c 268 g 204 t 4 others  
ORIGIN

alignment\_scores:  
Quality: 1614.00 Length: 334  
Ratio: 5.091 Gaps: 11  
Percent Similarity: 94.910 Percent Identity: 92.814  
  
alignment\_block:  
US-10-044-807-2 x BM476141 ..  
  
Align seg 1/1 to: BM476141 from: 1 to: 1009  
  
270 ProLeuThrAlaAspPheIleValIysIleArgAsnSerGlySerAlaAs 286  
|||||  
26 CCACCTACAGCAGATTCATTTGTCAGAGTCGTAACTCGGGCTCCGCTGA 75  
|||||  
286 pSerThrValGlnPheIlePheTyrGlnProIleIleHisArgTrpArgG 303  
|||||  
76 CAGTACAGTCCAGTTCATCTCTATCAACCCATCATCCACGATGAGGAG 125  
|||||  
303 luThrAspPhePheProCysSerAlaThrCysGlyGlyTyrGlnLeu 319  
|||||  
126 AGACGGATTTCTTCCTTGCTCAGACCTGTGGAGGAGGTATCAGCTG 175  
|||||

```

320 ThrSerAlaGluCysTyrAspLeuArgSerAsnArgValValAlaAspG1 336
|||||
176 ACATCGGTGAGTGTCTAGATCTGAGGAGCAACCGTGTGTGTGTGACCA 225
|||||
336 nTyrCysHisTyrTyrProGluAsnIleLeuProLysProLysLeuGlnG 353
|||||
226 ATACTGTACATATACCCAGAGACATCAACCCAAACCAAGCTTCAGG 275
|||||
353 luCysAsnLeuAspProCysProAlaSerAspGlyTyrLysGlnIleMet 369
|||||
276 AGTGCAACTGTGATCTTGTGTCCAGCAGTGCAGGATACCAAGCAGATCATG 325
|||||
370 ProTyrAspLeuTyrHisProLeuProArgTyrProGluAlaThrProTyrTh 386
|||||
326 CCTTATGACCTTACCATCCCTTCTCGGTGGAGGCCACCCCATGGAC 375
|||||
386 rAlaCysSerSerCysGlyGlyGlyIleGlnSerArgAlaValSerC 403
|||||
376 CGCGTGTCTCTCCCTCGTGTGGGGGGGATCCAGAGCCGGCAGTTTCTCT 425
|||||
403 ysValGluGluAspIleGlnGlyHisValThrSerValGluGluThrLys 419
|||||
426 GTGTGGAGGAGACATCCAGGGGATGTCATCTCAGTGGAAAGTGGAAA 475
|||||
420 CysMetTyrThrProLysMetProIleAlaGlnProCysAsnIlePheAs 436
|||||
476 TGCATGTACACCCCTAGATGCCATGCCGAGGCCCTGCACACTTTTGA 525
|||||
436 pCysProLysTrpLeuAlaGlnGluTrpSerProCysThrValThrCysG 453
|||||
526 CTGCCCTAAATGGTGGACAGAGTGGTCTCCGTCGCACATGCATCATGTG 575
|||||
453 lyGlnGlyLeuArgTyrArgValValLeuCysIleAspHisArgGlyMet 469
|||||
576 GCCAGGGCTCAGATACCGTGTGTCTCTCATCGACCATCGAGGAATG 625
|||||
470 HisThrGlyGlyCysSerProLysThrLysProHisIleLysGluGluCy 486
|||||
626 CACACAGAGGCTGTAGCCCAAAACAAAGCCCAACATAAAAGAGGAATG 675
|||||
486 sIleValProThrProCysTyrLysProLysGluLysLeuProValGluA 503
|||||
676 CATCGTACCCACTCCCTGCTATAAACCCAAAGAGAAACTTCCAGTCCGAGG 725
|||||
503 laLysLeuProTrpPheLysGlnAlaGlnGluLeuGluGluGlyAlaAla 519
|||||
726 CCAAGTTGCCATGTTCAAAACAGCTCAAGAGCTAGAAAGAGAGCTGCT 775
|||||
520 ValSerGluGluProSerPheIleProGluAlaTrpSerAlaCysThr.V 536
|||||
776 GTGTACAGAGGAGCCCTGTTTCATCCAGAGCCCTGGTCGGNCTGCACNAG 825
|||||
536 alThrCysGlyValGlyThrGlnValArg.IleValArgCysGln.ValL 552
|||||
826 TCACCTGTGTGTGGGACCCAGCTGGGAATAGTCAGTGCACGNGGC 875
|||||
552 euLeuSerPhe.SerGlnSer.ValAlaAspLeu.ProIleAspGluCys 567
|||||
876 TCCGTGTTCTTCTCTAGNCCCGTGGGTGACCTGGCCCTATTATTCACCGATG 925
|||||
568 ...GluGlyProLys.ProAlaSerGlnArg.AlaCysTyrAla.GlyPr 582
|||||
926 TGGAAAGGGCTTAGCCAGCATCCAGCTGGGCTGTATTATGAGGGGCC 975
|||||
582 oCys...SerGlyGluLeuProGluPheAsnPro 592
|||||
976 ATGCCACGGGGGAAAAATCTCTGGAGTCCACCCC 1009
|||||

```

seq\_name: gb\_est2:BE888902

seq\_documentation\_block:

LOCUS BE888902 850 bp mRNA linear EST 20-OCT-2000  
 DEFINITION 601513828F1 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3915294 5',

# ACCESSION VERSION KEYWORDS SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

mRNA sequence.

BE888902

BE888902.1 GI:10345670

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgc.nci.nih.gov/

1 (bases 1 to 850)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLML at:

http://image.llnl.gov

Plate: LLAM9738 row: g column: 07

High quality sequence stop: 742.

FEATURES

source

1..850

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3915294"

/clone\_lib="NIH\_MGC\_71"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb."

BASE COUNT 202 a 219 c 228 g 201 t

ORIGIN

alignment\_scores:

Quality: 1175.50 Length: 269

Ratio: 4.798 Gaps: 5

Percent Similarity: 91.078 Percent Identity: 86.617

alignment\_block:

US-10-044-807-2 x BE888902 ..

Align seg 1/1 to: BE888902 from: 1 to: 850

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuPheLeuAl 17

|||||

44 ATGGAATGTCGCGGGCAACTCTGCGACACTGCTCTCTTTCTGGC 93

|||||

17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34

|||||

94 TTTCTGCTCTGAGTTCCAGACCGCAGCTCCGAGGAGGACCGGAGC 143

|||||

34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50

|||||

144 GCCTATGGATGCTCGGGCCCATGGAGTGAATGCTCAGCACCCTGCGGG 193

|||||

51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG1 67

|||||

194 GGTGGGCTCTCTACTCTGAGGCGCTGCTGAGCAGCAGAGCTGTA 243

|||||

67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84

|||||

244 AGGAAGAAATATCCGATACAGACATGAGTAATGTGACTGCCACCAG 293

|||||

84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspValLys 100

|||||

294 AAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATAATGATGTCAG 343

|||||

101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117

|||||

seq\_name: qb\_est1:BB633755

seq documentation block.

LOCUS	EST 26-OCT-2001
LOCUS B8633755	linear mRNA
DEFINITION	622 bp
	RIKEN full-length enriched, adult male spinal cord Mus
	musculus cdna A330077F03 5', mRNA sequence.

ACCESSION	BB633755
VERSION	BB633755.1 GI:16470129

**KEYWORDS** EST.

SOURCE house mouse.

ORGANISM *Mus musculus*

Eukaryota; Me

Mammalia; Eut

## REFERENCE

**AUTHORS** Arakawa, T., C

, Hiramoto, K.

, M., Koya, S.,

Okazaki, Y., C

,D., Shibata,

Tagami, M., Tagami, M.

**Muramatsu, M.**

FILE  
TOURNAY  
RIKEN MOUSE  
RIKEN MOUSE

**JOURNAL**  
**COMMENT**  
**Unpublished**  
**Contact**

COMMENT: YOSHI

Laboratory for

The Institute of Sciences Centre

The Institute  
1-7-77

Tel.: 81-45-50

Tel: 81-43-50  
Fax: 81-45-50

06-64-79 : yrbj

[illegible]

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_xref="A33007F03"
/clone_lib="RIKEN full-length enriched, adult male spinal
cord"
/sex="male"
/tissue_type="spinal cord"
/dev_stage="adult"

```

/lab\_host="DH10B"  
/note="Site-1: Sali; Site-2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATTCGACGTCTTTTGGTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGACGTCTTAATTAAATCCCCCCCCCCC 3']. cDNA was cleaved with xhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda

BASE COUNT	151 a	157 c	172 g	142 t
ORIGIN				

alignment_scores:		
Quality:	1102.00	Length: 207
Ratio:	5.350	Gaps: 0
Percent Similarity:	99.517	Percent Identity: 96.618

alignment\_block:  
US-10-044-807-2

Align seg 1/1 to: BB633755 from: 1 to: 622

13 LeuLeuPheLeuAlaPheLeuLeuLeuSerSerArgThrAlaAargSerGI 29  
|||||  
2 CTGCTCGTTCCTGGCTTTCCTACTCTGAGTTCGAGGACTGCACGCTCAGA 51

```
29 uGluAspArgAspGlyLeuTrpAspAlaTrpGlyProTrpSerGluCys 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 GGAAGACCGTGAAGCGCTCTGGGATGCTGGGGCCCTGGAGCGAGTGT 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46 erArgThrCysGlyGlyAlaSerTyrSerLeuArgArgCysLeuSer 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 CCAGAACCTGTGTGGGGTGCCTCTTATTCCTGAGACGCTGCCTGAGC 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 SerLysSerCysGluGlyArgAsnIleArgTyrArgThrCysSerAsnVa 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 AGCAGAGCTGTGAAGGAGAAATATTCGATATAGACATGCAGTATGT 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
79 lAspCysProGluAlaGlyAspPheArgAlaGlnGlnCysSerAlaH 96
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
202 GGACTGCCACCACGAGCGGTGATTTCCGAGCTCAACAGTGTCTGCTC 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 isAsnAspValLysHisHisClyGlnPheTyrGluTrpLeuProValSer 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
252 ATAACGAGCTCAAGTACCATGGACAGCTTTATGAATGGCTTCCTGTATCT 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 AsnAspProAspAsnProCysSerLeuLysCysGlnAlaLysGlyThrTh 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
302 AATGACCCCGACATCATCTCACTCAAGTCCCAACCAAGGAACACAG 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
129 rLeuValValGluLeuAlaProLysValLeuAspGlyThrArgCysTyr 146
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
352 CCTGTTGTGGAACCTAGCACCGCAAGGCTTATAGATGGTACGGGATCTATA 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 hrGluSerLeuAspMetCysIleSerGlyLeuCysGlnIleValGlyCys 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
402 CAGAACTCACTGGATATGTGCATCAGCGGCTATGCCAAATTTGTGGCTGT 451
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 AspHisGlnLeuGlySerThrValLysGluAspAsnCysGlyValCysAs 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
452 GATCACCAGCTGGGAGCAGCTCTCAAGAGGATAACTGTGGAGTGGCA 501
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 nGlyAspGlySerThrCysArgLeuValArgGlyGlnTyrLysSerGlnL 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
502 CGGAGATGGCTCAACGTCGCCGCTGGTCCGAGGCGCAGTATAAATCTCAGC 551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 euSerAlaThrLysSerAspAspThrValValAlaIleProTyrGlySer 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 TCTCTGCAGTAATTCGATGACACCGTGGTGGCCATTCCTTACGGCAGC 601
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 ArgHisIleArgLeuValLeu 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
602 AGACACATCCGCCCTGTCTTA 622
```

seq\_name: gb\_estl:AI917724

```
seq_documentation_block:
LOCUS      AI917724                568 bp    mRNA    linear    EST 14-DEC-1999
DEFINITION ttlic08.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2240462 3'
similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE ; mRNA sequence.
ACCESSION  AI917724
VERSION    AI917724.1 GI:5637579
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 568)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapb-1@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
```

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 922 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 437.  
Location/Qualifiers

FEATURES  
source  
1..568  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2240462"  
/clone\_lib="NCI\_CGAP\_G66"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site.1: Not 1; Site.2: Eco RI; Plasmid DNA  
from the normalized library NCI\_CGAP\_G64 was prepared, and  
ss circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneIDs  
1257096-1258631, 1469064-1470983, and 1475592-1476743).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 120 a 157 c 169 g 121 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 1044.00 Length: 189  
Ratio: 5.583 Gaps: 0  
Percent Similarity: 98.942 Percent Identity: 98.413  
alignment\_block:  
US-10-044-807-2 x AI917724/rev ..  
Align seg 1/1 to reverse of: AI917724 from: 1 to: 568

```
666 CysProAlaArgTrpGluIleGlyLysTrpSerProCysSerLeuThrCys 682
568 TGCCCGACCAAGTGGGANATTGGCAAGTGGAGTCCATGTAGTCTCACATG 519
682 sclyValGlyLeuGlnThrArgAspValPheCysSerHisLeuLeuSerA 699
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
518 TCGCGTCCGCTTACAGACCAGACGCTCTCTCGAGCCACCTGCTGTCCA 469
699 rgGluMetAsnGluThrValIleLeuAlaAspGluLeuCysArgGlnPro 715
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
468 GAGAGATGAATGAACAGATCATCTGCTGGCTGATGAGTGTGCGCCAGCCC 419
716 LysProSerThrValGlnAlaCysAsnArgPheAsnCysProProAlaTr 732
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 AAGCCCGACGCGTGCAGCTTGTAAACCGCTTTAATTGCCCGCCAGCCG 369
732 pTyrProAlaGlnTrpGlnProCysSerArgThrCysGlyGlyValG 749
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
368 GTACCCCTGCACAGTGGCAGCCGTGTCCAGAACGTTGCGGGGGGTTC 319
749 lnLysArgGluValLeuCysLysLysGlnArgMetAlaAspGlySerPhe 765
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 AGAAACGTTGAGCTTCTTTCAGACGAGCGCATGGCTGATGGCAGCTCC 269
766 GluLeuProGluThrPheCysSerAlaSerLysProAlaCysGlnGlnAl 782
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 GAGCTCTCTGAGACCTCTCTTCAGCTTCAAAACCTGCTGCCAGCAAGC 219
782 cCysLysLysAspAspCysProSerGluTrpLeuLeuSerAspTrpThr 799
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 ATGCAGAAAGATGACTGTCGCCAGCGAGTGGCTTCTCAGACTGGACAG 169
799 luCysSerThrSerCysGlyGluGlyThrGlnThrArgSerAlaIleCys 815
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 AGTGTTCACAAAGCTGCGGGGAAGGCCAGACTCGAAGCGCATTTGC 119
```



816 ArgLysMetLeuLysThrGlyLeuSerThrValValAsnSerThrLeuCy 832  
 |||||  
 118 CGAAAGATGCTGAAACCGGCTCTCAACGGTGTGCAATTCACCGCTGTG 69  
 |||||  
 832 sProProLeuProPheSerSerIleArgProCysMetLeuAlaThrC 849  
 |||||  
 68 CCGGCCCTGCTTCCTTCTTCTCAATCAGGCCCTGTATGCTGGCAACT 19  
 |||||  
 849 ysAlaArgProGlyArg 854  
 |||||  
 18 GTGCAAGGCACGGCGCG 2

seq\_name: gb\_htc:AK020115

seq\_documentation\_block:

LOCUS AK020115 979 bp mRNA linear HTC 19-JAN-2002  
 DEFINITION Mus musculus 12 days embryo male wolffian duct includes surrounding region cDNA, RIKEN full-length enriched library, clone:6720426809:Thrombospondin type 1 domain containing protein, full insert sequence.

#### ACCESSION

VERSION AK020115.1 GI:12860599

KEYWORDS HTC; CAP trapper.

#### SOURCE

Mus musculus (strain:C57BL/6J) 12 days embryo male wolffian duct includes surrounding region cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:6720426809.

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; 1 (sites)

#### REFERENCE

##### AUTHORS

Carinci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

##### MEDLINE

99279253

##### PUBMED

10349636

#### REFERENCE

##### AUTHORS

Carinci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

##### MEDLINE

20499374

##### PUBMED

11042159

#### REFERENCE

##### AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format

JOURNAL sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

##### MEDLINE

20530913

##### PUBMED

11076861

#### REFERENCE

##### AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

#### REFERENCE

##### AUTHORS

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, X., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

#### TITLE

#### JOURNAL

Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

#### COMMENT

Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAAGGATCCAGAGCTCTTTTCTTTTCTTTTCTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCCAGTTAAATTAAATATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

#### FEATURES

##### Location/Qualifiers

1..979

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="MGI:1911603"

/db\_xref="taxon:10090"

/clone="6720426809"

/sex="male"

/tissue\_type="wolffian duct includes surrounding region"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="12 days embryo"

85..663

/note="Thrombospondin type 1 domain containing protein

putative"

data source: Pfam, source key: PF00090, evidence: ISS

/codon\_start=1

/protein\_id="BAB32000.1"

/db\_xref="GI:12860600"

/translation="MECCRAAPGTPLLVLAFLLLSSRTARSEEDREGLDWAGPWE

CSRTGGGASYSRLRCLSSKCEGRNIRYTCNVDPPEAGDFRAQQCSAHNDVKYH

GQLYEWLPVSDPDNPCLSKQAKGTSLVLEAPKVLVYGTGCTVTSIDMCISLUCIV

GCDHQIGSTVKKDKNCVCNGDGTCTCGVPRAV"

BASE COUNT 252 a 226 c 265 g 236 t

ORIGIN

#### alignment\_scores:

Quality: 1037.00 Length: 201

Ratio: 5.291 Gaps: 1

Percent Similarity: 97.512 Percent Identity: 94.030

#### alignment\_block:

US-10-044-807-2 x AK020115 ..

Align seg 1/1 to: AK020115 from: 1 to: 979

1 MetGluCysCysArgArqAlaThrProGlyThrLeuLeuLeuPheLeuAl 17  
 85 ATGGAATGCTGCGCTCGCGCAGCTCTCTGGCACACCGCTCTCGTTCTGCG 134  
 17 apheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34  
 135 TTTCCTACTCTCTGAGTCCAGGACTGCGCTCAGAGGAAGACCGTGAAG 184  
 34 lyLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50  
 |||||



```

34  lYLeuTrpAspAlaTrpGlyProTrpSerCluCysSerArgThrCysGly 50
|||||
205 GCCTCGGAGTCTGGGGCCCTGGAGGAGTCTCCAGAACCTGTGGT 254
51  GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLyssSerCysG1 67
|||||
255 GGGGGTGCCTCTATTCCTTGAGACCTGCTGAGCAGCAAGAGCTGTGA 304
67  uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
305 AGGGAGAAATATTCGATATAGAACATGCAGTAATGTGGACTGCCACAC 354
84  luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspVallys 100
|||||
355 AAGCAGGTGATTCGAGCTCAACAGTGTCTGCTCATACGACGTCACAG 404
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
|||||
405 TACCATGGACAGCTTATGATGGCTTCTGTATCTAATGACCCGACAA 454
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
455 TCCATGCTCACTCAAGTGCACAAAGCAAGAACAGCCCTGGTGTGGAAC 504
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
505 TACGACCCAGAGCTTATAGATGATCGCATGTATATACAGAAATCACTGG 554
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167
|||||
555 ATGTGCATCAGGGCTATGCCAAATGTGTGCTGTATCACCAGCTGGG 604
167 ySerThrVallysGluAspAsnCysGlyValCysAsnGlyAspGlySerT 184
|||||
605 GAGCACTGTCAAGAAGATAACTGTGGAATGTCAACGAGATGGCTCAA 654
184 hrCys 185
|||||
655 CGTGC 659

```

seq\_name: gb\_est2:BI523618

seq\_documentation\_block:  
LOCUS BI523618 581 bp mRNA linear EST 29-AUG-2001  
DEFINITION 603051757F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5201199 5',  
mRNA sequence.

ACCESSION BI523618

VERSION BI523618.1 GI:15348410

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC

TITLE NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11504 row: b column: 16

High quality sequence stop: 577.

Location/Qualifiers

1..581

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5201199"

FEATURES

source

seq\_name: gb\_est1:AI342006

seq\_documentation\_block:

```

/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/Note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026.. Note:
this is a NIH_MGC Library."

```

BASE COUNT 140 a 151 c 159 g 131 t  
ORIGIN

alignment\_scores:

Quality: 958.00 Length: 172

Ratio: 5.570

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-10-044-807-2 x BI523618

Align seg 1/1 to: BI523618 from: 1 to: 581

```

1 MetGluCysCysArgArgAlaThrProGlyThrLeuLeuPheLeuAl 17
|||||
66 ATGGGAATGCTGGCGTGGGCAACTCTGGCACACTGCTCTTTCTGGC 115
17 aPheLeuLeuLeuSerSerArgThrAlaArgSerGluGluAspArgAspG 34
|||||
116 TTCTCTGCTCTGAGTCCAGGACCCACGCTCCGAGGAGGACCGGAGC 165
34 lYLeuTrpAspAlaTrpGlyProTrpSerGluCysSerArgThrCysGly 50
|||||
166 GCCTATGGGATGCCTGGGGCCCATGGAGTGAATGCTCAGCACCTCGGG 215
51 GlyGlyAlaSerTyrSerLeuArgArgCysLeuSerSerLysSerCysG1 67
|||||
216 GGAGGGGGCTCTACTCTCTGAGGGCTGCCCTGAGCAGCAAGAGCTGTC 265
67 uGlyArgAsnIleArgTyrArgThrCysSerAsnValAspCysProProG 84
|||||
266 AGGAGAATAATCCCATACAGAACATGCAGTAATGTGGACTGCCACAC 315
84 luAlaGlyAspPheArgAlaGlnGlnCysSerAlaHisAsnAspVallys 100
|||||
316 AAGCAGGTGATTCGAGCTCAGCAATGCTCAGCTCATATGATGTCAG 365
101 HisHisGlyGlnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117
|||||
366 CACCATGGCCAGTTTATGAATGGCTTCTGTGCTAATGACCCCTGACAA 415
117 nProCysSerLeuLysCysGlnAlaLysGlyThrThrLeuValValGluL 134
|||||
416 CCCATGTTCACTCAAGTCCCAAGCCAAAGGAACAACCCCTGGTGTGTA 465
134 euAlaProLysValLeuAspGlyThrArgCysTyrThrGluSerLeuAsp 150
|||||
466 TAGCACCTTAAGTCTTAGATGGTAGCGCTTGTATATACAGAACTTTGG 515
151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167
|||||
516 ATGTGCATFCAGTGGTTATGCCAAATTTTGGCTCCGATCACCAGCTGG 565
167 ySerThrVallysGlu 172
|||||
566 AAGCACCGTCAAGGAA 581

```

```

LOCUS       AI342006               522 bp      mRNA      linear      EST 29-DEC-1998
DEFINITION   qs98d10.x1 NCI-CGAP_G4 Homo sapiens cDNA clone IMAGE:1946131 3'
              similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE 1, mRNA sequence.
ACCESSION   AI342006
VERSION     AI342006.1 GI:4078933
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 522)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapb@email.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -400P from Gibco
            High quality sequence stop: 482.

FEATURES             Location/Qualifiers
     source           1..522
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1946131"
                     /clone_lib="NCI-CGAP_G4"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT73
                     vector. Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonalodo."
BASE COUNT      106 a 139 c 166 g 110 t 1 others
ORIGIN
alignment_scores:
  Quality: 956.00      Length: 173
  Ratio: 5.558        Gaps: 0
Percent Similarity: 99.422 Percent Identity: 99.422

alignment_block:
US-10-044-807-2 x AI342006/rev ..
Align seg 1/1 to reverse of: AI342006 from: 1 to: 522

693 CysSerHisLeuLeuSerArgGluMetAsnGluThrValIleLeuAlaAs 709
|||||
521 TCGACCCACCTGCTTCCAGACAGATGAATGAACAGTCATCCTGGCTGA 472
|||||
709 pGluLeuCysArgGlnProLysProSerThrValGlnAlaCysAsnArgp 726
|||||
471 TGAGCTGTGTGCGCCAGCNCNCAAGCCAGCAGCGGTGCAAGCTTGTAAACCGCT 422
|||||
726 heAsnCysProProAlaTrpTyrProAlaGlnTrpGlnProCysSerArg 742
|||||
421 TTAATTTGCCCCCAGCTGTTACCTGACAGATGGCGAGCGGTTCACGA 372
|||||
743 ThrCysGlyGlyValGlnLysArgGluValLeuCysLysGlnArgMe 759
|||||
371 ACCTGTGGCGGGGTGTTTCAGAAACCTGAGGTCTTTTGAAGCAGCGCAT 322

```

```

759 tAlaAspGlySerPheLeuGluLeuProGluThrPheCysSerAlaSerL 776
|||||
321 GGCTGATGGCAGCTTCCCTGGAGCTTCTTGAGACCTTCGTTCAGCTCAA 272
|||||
776 ySProAlaCysGlnAlaCysLysLysAspCysProSerGluTrp 792
|||||
271 AACCTGCTGCCAGCAGCATGCAAGAAAGATGACTGTCCCGAGGTGG 222
|||||
793 LeuLeuSerAspTrpThrGluCysSerThrSerCysGlyGluGlyThrGl 809
|||||
221 CTCTCTCAGACTGGACAGAGTGTTCACAAAGCTGCGGGGAAGGCCCA 172
|||||
809 nThrArgSerAlaIleCysArgLysMetLeuLysThrGlyLeuSerThrV 826
|||||
171 GACTCGAAGCGCAFTTCCGAAAGATGCTGAAAACCGCCCTCTCAACGG 122
|||||
826 alValAsnSerThrLeuCysProProLeuProPheSerSerSerIleArg 842
|||||
121 TTGTCAATTCCACCTGTGCGCGCCCTGCCTTCTCTTCCCTCCATCAGG 72
|||||
843 ProCysMetLeuAlaThrCysAlaArgProGlyArgProSerThrLysHi 859
|||||
71 CCCTGTATGCTGGCAACCTGTGCAAGCGCGCGCGGCCATCCACGAAGCA 22
|||||
859 sSerProHisIleAlaAla 865
|||||
21 CAGCCCGCACATCGCGGCC 3
|||||
seq_name: gb_estl:BB643318

seq_documentation_block:
LOCUS       BB643318               700 bp      mRNA      linear      EST 26-OCT-2001
DEFINITION   BB643318 RIKEN full-length enriched, 9.5 days embryo parthenogenote
              Mus musculus cDNA clone B130031C01 5', mRNA sequence.
ACCESSION   BB643318
VERSION     BB643318.1 GI:16478052
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 700)
AUTHORS    Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
            Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sasaki
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
            Unpublished (2001)
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gs.riken.go.jp,
            URL: http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            waki,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

```

290 AAGCAGTGATTTCCGASCTCAACAGTGTCTGCTATACGACGTCACG 339  
101 HisHisGlyClnPheTyrGluTrpLeuProValSerAsnAspProAspAs 117  
340 TACCATTGGACAGCTTTTAAGATGGCTTCCGTGATCTATCTAATGACCCGACAA 389

117 nProCysSerLeuIysCysGlnAlaIysGlyThrLeuValValGluL 134  
 |||||  
 390 TCCATGCTCTACTCAAGTGCCAAAGCAAAAGGAACACGCGCGCTGTGGAAC 439  
 |||||

134 euAlaProIysValLeuAspGlyThrArgCystYrThrGluSerLeuAsp 150  
 440 TAGCACCGAAGGCTTAGATGTTACGGGATGCTATACAGAATCACTGGAT 489  
 151 MetCysIleSerGlyLeuCysGlnIleValGlyCysAspHisGlnLeuG1 167  
 490 ATGTGCATCAGCGGCGCTATGCCAAATGTTGGCTGTGATCACCCTGGG 539  
 167 ySerThrValLysGluAspAsnCysGlyValCysAsnGlyAspGlySer 183  
 540 GAGCACTGTCAAGAGAGGATTAACGTGGGAGTGGGCAACGGAAATGGCTCA 588

540 GAGCAGCTGTCAAGAGGATAACTGGGGAGTGGGCAACGAAATGGCTCA 588  
name: gb\_est1.BB478282  
documentation\_block:  
BB478282 619 bp mRNA linear  
BB478282 RIKEN full-length enriched, 13 days embryo  
musculus cdna clone D330029I02.3', mRNA sequence.

ENCE  
HORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki,  
1 (bases 1 to 619)

Zukavota; Metazoa; Chordata; Craniala; Vertebrata;

Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,  
 M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
 D., Shibata, K., Shingawa, A., Shiraki, T., Soqabe, Y.

Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, Y., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)  
On Jul 23, 2000 this sequence version replaced gi:9399  
Contact: Yoshihide Havaishizaki

Laboratory for Genome Exploration Research Group, RIKEN  
Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKENS)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 225  
Tel: 81-45-503-9222

Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shinozaki, K., and Taniguchi, M. 2000. The rice genome, <http://www.genome.jp/oryza.htm>.  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shinozaki, K., Taniguchi, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizawa, Y. 2000. The full-length complementary DNAs of the human genome. *Nature* 407:31-36.

Normalization and subtraction of cap-trapper-selected  
prepare full-length cDNA libraries for rapid discovery  
genes. *Genome Res.* 10(10): 1617-1620 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, S. Yoneda, Y. (2007) 1041 1030 (2000)

RIKEN integrated sequence analysis (RISA) system--38

sequencing pipeline with 384 multipillarary sequencers  
10 (11), 1757-1771 (2000)  
Konno, H. Fukunishi, Y. Shibata, Y. Tachibana, T.

Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA

encyclopedia; real-time sequence clustering for constant  
nonredundant cDNA library. Genome Res. 11 (2), 281-













151 MetCysIleSerGlyLeuCys 157  
|||||  
530 ATGTGCAATCAGCGGCTATGC 550



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 02:24:52 ; Search time 25.66 Seconds  
(without alignments)  
1677.237 Million cell updates/sec

Title: US-10-044-807-2  
Perfect score: 9588  
Sequence: 1 MECCRRATPTLLFLAFL.....LKLCQLSQKRCRCGCGKA 1762

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/6C\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/6D\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	373	3.9	788	2	US-08-918-914-4
2	359.5	3.7	441	3	US-08-985-526-3
3	311.5	3.2	1447	4	US-09-041-886-25
4	311.5	3.2	1447	5	PCT-US94-05277-2
5	307.5	3.2	1260	4	US-08-506-2968-21
6	304.5	3.2	469	1	US-08-313-2888-15
7	299	3.1	1253	4	US-08-506-2968-14
8	299	3.1	1268	4	US-08-506-2968-28
9	294	3.1	612	2	US-08-752-307B-11
10	294	3.0	1297	4	US-09-540-245A-17
11	290.5	3.0	615	2	US-08-752-307B-9
12	279.5	2.9	1651	4	US-09-540-245A-18
13	277.5	2.9	802	1	US-07-862-021B-12
14	277.5	2.9	802	1	US-07-862-021B-12
15	277.5	2.9	802	1	US-07-862-021B-12
16	272.5	2.8	607	2	PCT-US93-03164-12
17	264	2.8	807	1	US-08-752-307B-12
18	264	2.8	807	1	US-07-862-021B-10
19	264	2.8	807	1	US-08-313-288B-10
20	261	2.7	869	1	PCT-US93-03164-10
21	261	2.7	869	1	US-08-374-834-16
22	257.5	2.7	630	2	US-08-644-271-29
23	252.5	2.6	611	2	US-08-752-307B-14
24	247.5	2.6	868	1	US-08-752-307B-10
25	247.5	2.6	868	1	US-08-374-834-1
26	246.5	2.6	596	2	US-08-644-271-1
27	246	2.6	946	5	US-08-752-307B-13
					PCT-US95-08493-13
					Sequence 4, Appl
					Sequence 3, Appl
					Sequence 25, Appl
					Sequence 2, Appl
					Sequence 21, Appl
					Sequence 15, Appl
					Sequence 14, Appl
					Sequence 28, Appl
					Sequence 11, Appl
					Sequence 17, Appl
					Sequence 9, Appl
					Sequence 18, Appl
					Sequence 12, Appl
					Sequence 12, Appl
					Sequence 12, Appl
					Sequence 10, Appl
					Sequence 10, Appl
					Sequence 16, Appl
					Sequence 29, Appl
					Sequence 14, Appl
					Sequence 10, Appl
					Sequence 1, Appl
					Sequence 13, Appl

ALIGNMENTS

RESULT 1  
US-08-918-914-4  
; Sequence 4, Application US/08918914  
; Patent No. 5876963  
; GENERAL INFORMATION:  
; APPLICANT: Mitchell, Peter  
; APPLICANT: Hutchinson, Nancy  
; APPLICANT: Lawton, Michael  
; APPLICANT: Magna, Holly  
; APPLICANT: Yocum, Sue  
; APPLICANT: Murry, Lynn E.  
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/918,914  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0369  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 788 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1070094  
US-08-918-914-4

Sequence 15, Appl  
Sequence 19, Appl  
Sequence 21, Appl  
Sequence 15, Appl  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 16, Appl  
Sequence 20, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 3, Appl

```
Query Match 3.9%; Score 373; DB 2; Length 788;
Best Local Similarity 25.3%; Pred. No. 2.3e-20;
Matches 133; Conservative 49; Mismatches 156; Indels 188; Gaps 29;

QY 341 YPENIKPKKLOECNLDPCPASDGGYKQIMP-----YDLIYPLPWEATPTWATSSCGGGTQ 397
DB 429 YTRYRPAP-----PPPPACDGGCVNPPVVGWHD---W--SDWSTCSTCGDGAK 476

QY 398 SRVSCVEDIOGHVTSVEEWKCMYTPKPIAOPCNIFDCPKWLAQ--EWSPTCTVTCGGGL 456
DB 477 SRRECESTNNCGADYETE-----PCNLGPCQWSEWCEWSTUSASCGSQ 522

QY 457 RYRVVICIDHRGHTGCGSPKTKPHKECIVPTCPYKPKLEPLVEAKLPWFKAQAELEE 516
DB 523 RETRFE-----HLGTNRCBGKDYSEQCSA--GPC-----PENSOWED--- 559

QY 517 GAVSEEPSFIPEAWSACTVCGVGTQVIRVRCQVLLSFSQSVDLPIDB--CEGPKPASQ 575
DB 560 -----WQCSVTCGGVAVRQRTC-----LGGVFGDHLCCQPK--TEQ 595

QY 576 RACYAGPCSEIPEFNPDETDGLFGGLQDFDELYDWEYEGFTKCSBSCGGGVQEAUVSC-- 634
DB 596 RACDGGPCSLWSB-----WQEWSTCSASCSGSMKRRQVQC 631

QY 635 LNKOTREPAENLCVTSRRPPQLLKSCLNDPCPARW--EIGKWSPCSLTCGVGLQTRDVFC 693
DB 632 FGTDCCQPNES-----QFCYGPCC--AEWTEWCEWSGCSKCGPGQRTTRGC 678

QY 694 SHLLSREMNETHVILADELCROPKESTV-----QACNRENCPPAWYPAQWPCSTCGGGVQ 749
DB 679 LGPNGQEA-----TCQGPSIETTLCEGSCCNWS---EW--CHWSMCDKECGGG-- 723

QY 750 KREVLCQRMADGSFLELPETFCASAKPACQACKDDCPSEWLLSDWTECST--SCGEG 807
DB 724 -----QVRYIEYMF-----RTGC-----EWSPCSTQLACEVG 750

QY 808 TOTRSALCRKMLTGLSTV-----VNSTLCPPPLPFSSIRPC 844
DB 751 VQSRROC--VGESGCHCIGLAESQOCRLGTCPPKP-----PC 788

RESULT 2
US-08-985-526-3
; Sequence 3, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-985-526-3

Query Match 3.7%; Score 359.5; DB 3; Length 441;
Best Local Similarity 25.6%; Pred. No. 1e-19;
Matches 138; Conservative 58; Mismatches 180; Indels 163; Gaps 29;

QY 324 CY-----DLRSNRVADQYC--HYYPENIKPKPKLOECNLDPC-----PASDGYKQIMPYDL 373
DB 18 CYHNGVQYRNNEEDTDSCTECHQNSVTICKVYSCPIPCSNATVPDGCPCRCWPSD-- 76

QY 374 YHPLPWEA--TPWTACSSCGGQIQSRVSCVEEDTQGHVTSVEEWKCMYTPKMPIAQ 431
DB 77 --SADDMGSPWSEWTSCTSCGNIQORGRSCDSLNNRCGSSVQ-----TRT 122

QY 432 CNIFDCPKWLAQE-----WSPCTVTCGGGLRYRVVLCIDHRGHTG-----GCSPKTK 479
DB 123 CHIQECDKRFKQDGGWHSWSPSSCVTCGDGVITRITLCLNSPSPQMNGKPCGEARETK 182

QY 480 PHKECIVPTCPYKPKLEPLVEAKL--PWFKAQAELEGAASEPFIPEAWSACTVTC 538
DB 183 ACKKDAC-----PINGGWGPW-----SPWDICSVTC 208

QY 539 GVGTOVIRVRCQVLLSFSQSVDLPIDBCEGPKPASQ---RACYAGPCSEIPEFNPDE 594
DB 209 GGGVQKR-----SRICVDSRMTE--ENKELANELRRPLCY----- 242

QY 595 TDGLFGGLQDFDELYDWEYEGFTKCSBSCGGGVQEAUVVSCLNKOTREPAENLCVTSRRP 654
DB 243 -----HNGVQ--YENNEEWTVDSCTECH-----CQNSVTIC--KKVSCPIMP--CSNATVP 287

QY 655 -----PQLLKSCLNDPCPARWEIGKWSPCSLTCGVGLQTRDVFCSHLLSREMNETHVILAD 709
DB 288 DGECCPCRCWPSDSADDDGWSPW--SEWTSCTSCGNIQORGRSCDSLNNRCGSSV--- 341

QY 710 ELCROPKESTVQACN--RENCPPAW--YPAQWOPCSTCGGVQKREVLCQK-----RMA 760
DB 342 ----QTRTCHIQECDKRFKQDGGWHSWSPSSCVTCGDGVITRITLCLNSPSPQMNGKPC 397

QY 761 DGSFLELPETFCASAKPACQACKDDCP-----SEWLLSDWTECSTCGEGTQTRSAI 814
DB 398 EG---EARET-----KACKKDACPINGGWGPW--SPWDICSVTCGGGVQKRSRL 441

RESULT 3
US-09-041-886-25
; Sequence 25, Application US/09041886
; Patent No. 6215872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Razizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-041-886-25

```

```

Query Match          3.2%; Score 311.5; DB 4; Length 1447;
Best Local Similarity 20.7%; Pred. No. 3.9e-15;
Matches 151; Conservative 89; Mismatches 215; Indels 275; Gaps 32;

```

```

QY 1148 SSTGDAGGSRPHRKPTILRKISAA---QQLSASEVVT-HLGTVALASGTLISVLLHCE 1203
Db 119 ASLGDGSGSIISR-----TAKVAVAGPLRFLSQTESVTAFMGDT-----VLLKCE 162
QY 1204 AIGHPRPTISWANGEE---VQFSRLLLPDSDLSQILAPVEADVGYTCNATNALGELTES 1260
Db 163 VIGEPMTIHWKQNOQLTPIPGDSRVVVLPSCALQISRLQPGDIGIYRCSARNP----- 217
QY 1261 VSTAVILACKPLVKTSRMVTINTEKPAVTVDIG-----STIKTVQGVNVTINLQC 1309
Db 218 ----ASSRTG-----NEAEVRILSDPGLRHQLYFLQRPNSVVAIEGDAVLECC 262
QY 1310 VAGVPEAEVTFWFRNKSCL---GSPHILHEGS-LLLTNVSSDGLYSCRAANLHGLTES 1365
Db 263 VSGYPPSPFTWLRGEVIOLRKSKYLLGGSNLLISNVTDDSGMVTCTVYTKNENISAS 322
QY 1366 TOLLILDPQVPTQLEDIRALLAATGPNIPLSPVLTSPILGTQLVDPGNSALLGCFIKGHPV 1425
Db 323 AELTVLVPPWF-----LNHPSNLXAYESMDIEFE-----CTVSGKPV 359
QY 1426 PNITWPHGQPIVATAGLTHHILAAQOILQVANILOGSGOGFESCLAQNPAGVLMQKASLV 1485
Db 360 PTVMNMKNGDVVIPS---DYFQIVGGSNLRILGVKSDGFGYQCVAEAGNAQTSQALI 416
QY 1486 IQDYVMSVDRLATCSASCGNRGVQOPRLCLLNSTEVNPAHCAGKVRPA-----VQPIACN 1541
Db 417 VP-----KPAIPSSSVLPSPAPRDVVPVLVS 441
QY 1542 RRDPSRMVMTSWSACTRSCGGGVOT-----RVVTCOKLKAS 1578
Db 442 ----SRFVRLSWRP-PAEAKGNIQITFTVFFSREGDNRREALNTQPGSLQTLVGNLKPE 495
QY 1579 GLST-----PVSDNMCQTQAKRP-----VDQACN---QQICVFEWAFSSWQ 1617
Db 496 AMYTFRVVAYNEWGGESSQPIKQVATQPELQVPGPVENLQAVSTSPSTSLITWEPPAY-- 553
QY 1618 CNGPCIGPHILAVOHRQVFCQ-----VQVYRFLCTEVSTKEQNIQVGGISYKLEGLKFKTEYSRLFLAYNRYGP 1637
Db 554 ANGP-----VQVYRFLCTEVSTKEQNIQVGGISYKLEGLKFKTEYSRLFLAYNRYGP 606
QY 1638 --TRDGIT-----LPSE-----QCSALPRPVSTONCWSEACSVHWRV-- 1672
Db 607 GYSTDDITWTLSDVPSAPPQNVSLVNVNSRIKYSWLPPLPPPGSTONGFTGTGKIRHKITT 666
QY 1673 -----SLWTLCTA--TCGNYGQFSRRVECVHARTNAKVAPEHLCSWGPGRPANWOR 1719
Db 667 RRGEMETLEPNLNLWFLTGLEKGSQYSFQ-----VSAMTVNGT-----GP-PSNWT 712

```

```

QY 1720 CNITPCENME 1729
Db 713 AE-TPENDLD 721
RESULT 4
PCT-US94-05277-2
; Sequence 2, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zablacky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-05277-2

```

```

Query Match          3.2%; Score 311.5; DB 5; Length 1447;
Best Local Similarity 20.7%; Pred. No. 3.9e-15;
Matches 151; Conservative 89; Mismatches 215; Indels 275; Gaps 32;

```

```

QY 1148 SSTGDAGGSRPHRKPTILRKISAA---QQLSASEVVT-HLGTVALASGTLISVLLHCE 1203
Db 119 ASLGDGSGSIISR-----TAKVAVAGPLRFLSQTESVTAFMGDT-----VLLKCE 162
QY 1204 AIGHPRPTISWANGEE---VQFSRLLLPDSDLSQILAPVEADVGYTCNATNALGELTES 1260
Db 163 VIGEPMTIHWKQNOQLTPIPGDSRVVVLPSCALQISRLQPGDIGIYRCSARNP----- 217
QY 1261 VSTAVILACKPLVKTSRMVTINTEKPAVTVDIG-----STIKTVQGVNVTINQC 1309
Db 218 ----ASSRTG-----NEAEVRILSDPGLRHQLYFLQRPNSVVAIEGDAVLECC 262
QY 1310 VAGVPEAEVTFWFRNKSCL---GSPHILHEGS-LLLTNVSSDGLYSCRAANLHGLTES 1365
Db 263 VSGYPPSPFTWLRGEVIOLRKSKYLLGGSNLLISNVTDDSGMVTCTVYTKNENISAS 322
QY 1366 TOLLILDPQVPTQLEDIRALLAATGPNIPLSPVLTSPILGTQLVDPGNSALLGCFIKGHPV 1425
Db 323 AELTVLVPPWF-----LNHPSNLXAYESMDIEFE-----CTVSGKPV 359

```





## RESULT 6

US-08-313-288B-15  
 ; Sequence 15, Application US/08313288B  
 ; Patent No. 5750502  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jessell, Thomas M. and Avihu Klar  
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/313,288B  
 ; FILING DATE: January 5, 1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
 ; TELEPHONE: (212) 278-0400  
 ; TELEFAX: (212) 391-0526  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 15:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 469 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; US-08-313-288B-15

Query Match 3.2%; Score 304.5; DB 1; Length 469;  
 Best Local Similarity 23.8%; Pred. No. 2.2e-15;  
 Matches 131; Conservative 42; Mismatches 164; Indels 213; Gaps 33;  
 QY 388 CSSSCGGGIGQSRVSCVEEDIQGHVTSVEEWKMYTP-----KMPAQPCNIFDCPKW- 440  
 Db 43 CKGLGGGV-----SVEDC-CLNTAFAYOKRSGGLQPCRC-----SPRWS 81  
 QY 441 LAQWSPCTVTCGGGLRYRVVLCIDHRMHTGGSPKTKPHKECIVPTPKYKPKLP 500  
 Db 82 LWSTWAPCSVTCSEGSOLRYRRCVGVNGQCSKVAPCT----- 119  
 QY 501 VEAKLPWFQAOLEEGAAVSEPSFIPE-----AWSACTVTCGVGTQVRLVRCQVL 552  
 Db 120 -----LEWQJQA-----CEDQCCCPENGSGWGWPWPCSVTCSKGTTRRRAC--- 163  
 QY 553 LSFSSQSVADLPIDCEGPKFA-----SORACVAGPCSGELPENPDETGLFGGLQ 603  
 Db 164 -----NHPAPKCGGCHCPQAOESACDQVVC-----PTHGAWATWGP----- 201  
 QY 604 DFDELYDMEYEGTKSESGGGVQEAUVSCLNKQTR-----EPAEENLCVTSRRPPQ 656  
 Db 202 -----WTPCSASCHGGPHEP-----KETRSKCSAPE-----SOKPPG 235  
 QY 657 LKASC-----NLDPCLPAREIGKW-----SPCLTCGVGLQTRDVFCSHLLSREM 702  
 Db 236 --KPCPLAYEQRRCRTGLPPCPVAGGWPVSPCPVTCGLGLQTMQRTCNHPVQH-- 291

QY 703 ETVILADELCRQPKFSTVQACN-RFNCP-----PWTYPAQWQPCSR-----TCG--GGV 748  
 Db 292 -----GGPFCAGDATRT-HICNTAVPCPVGDGWSW--GEWSPCIRRNMKSIQCEIPGQ 343  
 QY 749 QKREVLCKQRMADGSFLELPETFCASAKPACQACKKDDCP-----SEWLLSDWTEGTS 803  
 Db 344 QSRGRTCRGPKFDGH-----RCAGQQQDIRHCYSIQHCHPLKGSWSEW--STWGLCMPP 394  
 QY 804 CGSG-TQTSRAICRMLKTGLSTV-----VNSTLC-PPLPFSSSI-----RP 843  
 Db 395 CGPNPTRAQRQLCTPLLPKYPPTVSMVGEQCKNVTFWGRPLPRCEBLOGLQKLVWEKRP 454  
 QY 844 CM-LATCARP 852  
 Db 455 CLHVPACKDP 464  
 RESULT 7  
 US-08-506-296B-14  
 ; Sequence 14, Application US/08506296B  
 ; Patent No. 6313265  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Phillips, Greg  
 ; APPLICANT: Cunningham, Bruce A.  
 ; APPLICANT: Crossin, Kathryn L.  
 ; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES  
 ; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE  
 ; NUMBER OF SEQUENCES: 77  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: The Scripps Research Institute  
 ; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8  
 ; CITY: La Jolla  
 ; STATE: California  
 ; COUNTRY: U.S.  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/506,296B  
 ; FILING DATE: 24-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fitting, Thomas  
 ; REGISTRATION NUMBER: 34,163  
 ; REFERENCE/DOCKET NUMBER: TSRI 488.0  
 ; TELEPHONE: (619) 554-2937  
 ; TELEFAX: (619) 554-6312  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1253 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-506-296B-14

Query Match 3.1%; Score 299; DB 4; Length 1253;  
 Best Local Similarity 20.1%; Pred. No. 2.9e-14;  
 Matches 187; Conservative 110; Mismatches 345; Indels 288; Gaps 36;  
 QY 843 PCMLATCARPGRTSKHSPHIAARKVYIOTRRQRKLFHVVGGFAYLLPKTAVVLRCPAR 902  
 Db 16 PCLLIQI-----PEEYEGHHVMEPPVITEQSPR-----LVVPTDDISLKCEAS 60  
 QY 903 RVRKPLITWCKDQHLISSTHTVTVAPPGYLIHRLKPSDAGVYTCAGPAREHFVKLIG 962  
 Db 61 GKPEVOFRWTRDGVHFKPKPELGVTVY-----QSPHSGSFITGN--NSNFAQRFQ 110  
 QY 963 GNRKLVARPLSPREEVLAGRGKGPKEALQTHKHQNGIFNSGKAEKRGLAANPGSRD 1022

111 IYCFASNKLGITAMSHIRLMASGAPK-----137  
Db  
1023 DLVSRLLQGGWPGELLASWEADSNERTTSEEDGAEQV-----LLHLPTMTVTE 1074  
QY  
138 -----WPKETVKPVEVEGESVVLPCNPPPSAEPLRIYWMNSKILHIK-----QD 182  
Db  
1075 QRRLLDILGNLSQPELRDLYSKHLVAQAEIFRSHLEHODTLKPSERRTSPVTLSP 1134  
QY  
183 ERYTMGONGNLYFANVLTSDNHSDY-----ICHAFPGTRTII-----QKEPI----- 225  
Db  
1135 HKHVSFSSSLRTSSGDAGGRRRPHKPTILRKISAAQOLASAEVTHLQGVVALASG 1194  
QY  
226 -----DLRVKAT-----NSMIDRKPRLLFPTN-----SSSHLVALQGG----- 258  
Db  
1195 TLSVLLHCEAIGHPRPTISWARGNEEVQFSDRILLO-PDSSLQTLAPVEADVGYTCNAT 1253  
QY  
259 --PLVLECIAGFPTPTIKWLRPSGMP-ADRVTYQNHNTKTLQLLKYGEEDGGEYRCLAE 315  
Db  
1254 NALGYDSVSTAVTLAGPLVKTSMTVINTKPAVTVDIGSTIKTVQGVNNTINQVAGV 1313  
QY  
316 NSLGSARHAYVTEAPYW-----LHKP-----QSHLYGGETARLDCQVQGR 359  
Db  
1314 PEAETWFRNKSILGSP-----HLHFGSLLLLTNVSSSDQGLYSCRAANLHGELTE 1364  
QY  
360 PQEVTWRIN---GIPVEELAKDQYRIQKALILSNVQPSDTMTVQCEARNRHGLLLA 415  
Db  
1365 STQLLIDPP-----QVP-----TOLED----- 1382  
QY  
416 NAYIVVOLPAKILTDADNTYNAVQGSTAYILLKAPGAPVPSVQWLDDEGTVIQUERFF 475  
Db  
1383 -----TRAL-----LAATGNPLPSVLTSS-----PLGTQLVLDP-----GNSA 1414  
QY  
476 PYANGTIGIRLDQANDTGRYFCLAANDQNNVTIMANLKVADATQITQGRSTIEKKGSRV 535  
Db  
1415 LIGCPKIGHP--VPNITWFGGQPIVATGTHHILAAQIILQVANLGGSGGEGESCLAQ 1472  
QY  
536 TFTQASDPDSLSQPSITWKGDRDLQELGDSKYPFIEDGRV-THSLDYSDQCNYSVAS 594  
Db  
1473 NEAGVLMQKASLVIQDYWMSVDRLATCSACGNRGVQPRCLLNSTEVNPAHCAGKVR 1532  
QY  
595 TELDVESRAQLLVVSGPVPRLVLSLH-----LITQSVRVSWSPADH 641  
Db  
1533 PAVQPIAC-----NRRDCPSRW-----MWTWSACTRSCGGVQ-TRRVTCKLKASGI 1580  
QY  
642 NA-PIEKVDIEFEKKAPEKNWYSLGKVPQGTSTTLKLSYVHYVTRVTAINKYGPGE 699  
Db  
1581 STPVNSDMCTQVA---KRPVDTQACNQ---LCVWEAFSSWQCGNGPCIGPHLAYOHRQV 1634  
QY  
700 PSPVSETVVTPEAAPEKNVDVKGEGNETTNMTVITWKLPLRMDWNAPE-----QVQYRVQ 753  
Db  
1635 FCQTRDGITLPSEQSALPRPVSTQNCWSE 1664  
QY  
754 W-----RPOGRGPWQE 765  
Db

RESULT 8  
US-08-506-296B-28  
; Sequence 28, Application US/08506296B  
; Patent No. 6313265  
; GENERAL INFORMATION:  
; APPLICANT: Phillips, Greg  
; APPLICANT: Cunningham, Bruce A.  
; APPLICANT: Crossin, Kathryn L.  
; APPLICANT: Crossin, Kathryn L.  
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES  
; NUMBER OF SEQUENCES: 77  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute  
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: U.S.

ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/506,296B  
FILING DATE: 24-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: ISRI 488.0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 554-2937  
TELEFAX: (619) 554-6312  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1268 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-506-296B-28

Query Match 3.1%; Score 299; DB 4; Length 1268;  
Best Local Similarity 20.6%; Pred. No. 3e-14;  
Matches 179; Conservative 119; Mismatches 327; Indels 246; Gaps 37;

QY 891 PXTAVVLCPRARRVRKPLITWEKQGH--LISSTHTVAP-FGYLKHRLKPSDA---G 943  
Db 55 PRENIVIOCEAKGKPPSPFSTWRNGTHEDIDKDAQYTKPNSTGLVNMNGVKAAYEG 114  
QY 944 VYTCAGAPAREHIVKILGNGKLVARP-LSPSRSEEVLAGRKGPKKEALQTHKQNGIF 1002  
Db 115 VYQCTARNERGAII-----SNIVLRPSRL-----WTKEKLEPHVREG-- 155  
QY 1003 SNGSKAEKRGLANPGSRVDDLVSRLEGGWPGFELLASWEAQDSNERTTSEEDPGAQ 1062  
Db 156 -----DSLVNCRPPVGLPPTII-FW--MDNAFQL-----PQSER 188  
QY 1063 VLLHLPTMTVQRRLLDILGNLSQPELRDLYSKHLVAQAEIFRSHLEHODTLKLP 1122  
Db 189 V-----SQUNGDLIFSNOVQEDIRVDY-----ICYARFNITQI--- 223  
QY 1123 SERRTSPVTLSPKHVSGFSSSLRTSSTGDAGGSRPHKPTILRKISAAQOLASAEV 1182  
Db 224 --QKQKQPIIS-----VKVFSTKPTV-----ERPPVLL----- 247  
QY 1183 THLGQT---VALASGTLVLLHCEAIGHPRPTISWARGNEEVQFSDRILLOPDSLOILA 1239  
Db 248 TPMGSTSNKVELRGVNL--LLECIAAGLTPVIRWIKEGEGELPANRTFFENFKKTLKIID 305  
QY 1240 PVEADVGYTCNATNALGYDSVSIATVLACKPLVKTSMTVINTKPAVTVDIGSTIKTV 1299  
Db 306 VSEADSGNKKCTARNILGTHHIVISVTKAAPTWTAPRNLVLS----- 349  
QY 1300 QGVNVTINQVAGVPEAEVTFWRNKSILG-----SPHLLHEG-SLLLTNWSSSDQGLYSR 1354  
Db 350 PEGDGLICRANGNPKPSISWLTNGVPIALAPEDPSRKVDGDTIIFSAVQERSAVVQCN 409  
QY 1355 AANLHGELESTQLLIL-DPPQVPTQLEDIRALLAAGPNLPSVLTSLGTQLVLDPGNS 1413  
Db 410 ASNEYGYLLANAFVNLAEPRILTPANKLYQVIA-----DSP----- 447  
QY 1414 ALLGCPKIGHPVNIWTFHGGQPIVATGTHHILAAQIILQVANLGGSGGEGESCLAQ 1473  
Db 448 ALIDCATFGSPKPIEFWRGVKGSILRG--NEYVFDHNGTLEIPVAKDSTGTGTCVARN 505  
QY 1474 EAGVLMQKASLVIQDYWMSVDRLATCSACGNRGVQPRCLLNSTEVNPAHCAGKVR 1533  
Db 506 KLKGTQNEVQLEVKD-----PTMIKQPOYKVIQSAQAS-FECVTKHDP 549

QY 1534 AVOPIAC-----NRDCPSRWV-----TSWACTRSCGGVOTRVRTCQKLGASISTP 1583  
DB 550 TILPTVINLKDNNELPDDREFLWVGKDNLTIMNVTDKDDGTTCIVNTTLDSSVSASAVLT 609  
QY 1584 VSDMCTQV---AKRPVDTQACNQ-OLCVENAFSSWGCGPC1-----1623  
DB 610 VAAPPTPAILIARPNPDLDELTLGQLERSIELSWPGGEENNSPITNFVIEYEDGLHEPGV 669  
QY 1624 -----GPHLAVOHR-----OVFCQTRDGIITLPEQC-----SALP--RPVST 1658  
DB 670 WHYQTEVGSHTTIVOLKLSPPYNYFVAVNEIGRSQFSESEQYLTKSAPNDENPSNV 729  
QY 1659 QNCWEACSVHVRVSLWLTCTATCNGYFQFS 1689  
DB 730 QGIGSEPDML---VITWESLK-----GFQS 751

RESULT 9  
US-08-752-307B-11  
; Sequence 11, Application US/08752307B  
; Patent No. 5952171  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean A.  
; APPLICANT: Gearing, David P.  
; APPLICANT: Levinson, Douglas A.  
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES  
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,307B  
; FILING DATE: 19-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 09404/020001  
; TELEPHONE: 617-542-8906  
; TELEFAX: 617-542-5070  
; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 612 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-752-307B-11

Query Match 3.1%; Score 295; DB 2; Length 612;  
Best Local Similarity 21.2%; Pred. No. 1.9e-14;  
Matches 141; Conservative 98; Mismatches 243; Indels 182; Gaps 27;

QY 891 PXTAVVLRCPARRVRKPLTWKDKQH--LISSHTVYAP-FGYLKIHLKPSDA---G 943  
DB 55 PRENIVIOCEAKGKPPPSFSTRNGTHFDIDKDAQVTKPNSGTLVNMINGVKAFAYEG 114

QY 944 VTCSAGPAREHVIKLGNGNRKLVARP-LSPRSEEEVLAGRKGPKKALQTHKHQMGIF 1002  
DB 115 VYCTARNERGAAI-----SNNIVIRPSRPL-----WKEKLEPNHVRGS-- 155  
QY 1003 SNGSKAEKRGKLAANPGSRDYDDLYSRLLLEQGGWPGELLASWEAODSARNRTTSEDPGAEQ 1062  
DB 156 -----DSVLNCRPPVGLPPPII-FW--MDNAFQRL-----PQSER 188  
QY 1063 VILHLFTVMTQRRRDDILGNLSQOPEELRDLYSKHLAQLAQEIFRSHLEHODTLKLP 1122  
DB 189 V-----SQGLNGDLYFSNVQPEDTRVDY-----ICYARFNTQTI--- 223  
QY 1123 SERTTSPVTLSPHKHVSFGFSSSLRTSTGDAGGSRPRPKPILRKISAAQOLSAEVV 1182  
DB 224 ---CQKQPS-----VKVFSTKPV-----ERPPVLL----- 247  
QY 1183 THLGOT---VALASGTLVLLHCEAIGHPRPTISWARGEEVQFSDRILLOPDDSLQILA 1239  
DB 248 TPMGSTSNKVELRGNVL---LLECIAAGLTPVIRWKEGGELPANRTFFENFKTKLIIID 305  
QY 1240 PVEADVGYTCNATNALGYDSVSIATVLAKPLVKTRMTVINTEKPAVTVDIGSTIKTV 1299  
DB 306 VSEADSGNYKCTARTNLGSTHHVSVTVKAAPYWITAPRNLVS----- 349  
QY 1300 QGVNVTINCOVAGVPEAEVTFWRNKSGLG-----SPHLLHEG-SLLLTNVSSSDQGLYSCR 1354  
DB 350 PGEDTLCIRANGNPKPSISWTNGVPIAIPEDPSRKVDGTIIFSAVQERSASSAVYQCN 409  
QY 1355 AANLHGETESTQLLIL-DPPQVPTQLEDIRALLAATGPNLPSVLTSPLGTQLVLDPGNS 1413  
DB 410 ASNEYGYLLANAFVNVLAEPPIRLTPANKLQVIA-----DSP----- 447  
QY 1414 ALLGCPKIGHVPNTWPHGQPIVITATGLTHHILAAGLIQVANLSGSGCEFSCLAQN 1473  
DB 448 ALIDCAVFGSPKPELEWPRGVKGSILRG--NEYVPHDNGTLEIPVAKDSTGYTCVARN 505  
QY 1474 FAGVLMOKASLVIODYWWSVDRLATCSACGNRGVQOQPLRCLLNSTEVNPAHCAGKVRP 1533  
DB 506 KLGTONEVQLEVKD-----PTMIKQPYKVIVORSQAAS-FECVIKHPD 549  
QY 1534 AVQP 1537  
DB 550 TLIP 553

RESULT 10  
US-09-540-245A-17  
; Sequence 17, Application US/09540245A  
; Patent No. 6270984  
; GENERAL INFORMATION:  
; APPLICANT: Goodman, Corey  
; APPLICANT: Kid, Thomas  
; APPLICANT: Brosier, Katja  
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
; FILE REFERENCE: B98-031-3  
; CURRENT APPLICATION NUMBER: US/09/540,245A  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 60/065,544  
; PRIOR FILING DATE: 1997-11-14  
; PRIOR APPLICATION NUMBER: 60/081,057  
; PRIOR FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 1297  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-540-245A-17

Query Match 3.1%; Score 294; DB 4; Length 1297;  
Best Local Similarity 22.0%; Pred. No. 7.6e-14;



```

RESULT      12
US-09-540-245A-18
; Sequence 18, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1651
; TYPE: PRT
; ORGANISM: human
US-09-540-245A-18


Query Match          2.9%; Score 279.5; DB 4; Length 1651;
Best Local Similarity 20.6%; Pred. No. 1.5e+12;
Matches 155; Conservative 105; Mismatches 278; Indels 213; Gaps 30;

QY   1020 RYDDIVSRLLGEGGWPGELLAS-----WEAQDSAPRNTSTSEDPGAEE 1061
     :|::||:||:||||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db    62 ROEDPPRIVEH---PSDLVSKGPATLNCKAERGPTTIEWKGGSERVEDTKDPRS 118

QY   1062 QVLV---HLPTMTWTEQRRL-----DDLGNLSQPDELRLDYLSKH--- 1099
     ::||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db    119 RMLLPFGSLFELRVHGKRSPDEGVYYVCVARNYLGAVSNASLEVALLDPRNP 178

QY   1100 -LVLAQLAEIIF----RSHLEHQDTLIK--PSERTSPTVTLSPHKHVGFSSLRIT 1150
     :||| |:|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    179 VMVAAGEPAVMECOPPKGHPEPTISWKKGDSLDDKERIT-----IRGCKLMITYRK 232

QY   1151 GDAGG---GSRRPHRKPTILRKISAQAOLSASEVVTHLGTOTALASTGLSVLLHCRAIG 1206
     ||| | | :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    233 SDAGKYCVCGTNMVGERESEAVELTVLERFSFKRPNSLATV-----DTSABEFKCARG 287

QY   1207 HPRPTISKARNGEVQVSDRILLQPDOSLIQLAPEADVGYFTCNATNALGYDSVSIAVT 1266
     |||::|:|:~|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    288 DPVTVWRKKDGDELPLKS--RYEIHDDHTLKIRKVTAGDMGYSYCVAENMVYGKAFASATLT 346

QY   1267 LAGKP--LVKTISRMTVINTEKPAVTVDIGSTITVOGVNVTINCOVAGVPFAEVTFMRNK 1324
     :|:| ~|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    347 VQPPPHEFWVKPRQQV-----ALGRVTFOCEATGNPOPIAWFRRE-- 387

QY   1325 SKLGSPhHL-----HPGSLLLFNVSSTDGLYSCRAANLHELGETSQL 1369
     ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db    388 ---GSONLLFSYPQOSSRSFVSQTGDLITNVQRSDVGYICQLLNAGSIITKAYLE 444

QY   1370 ILDPDPQVPTOLEDIRALLAAATGNPLSVLTSGTGLOILDPNMSALNGCPIKGHPVNIT 1429
     :|:|::~|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    445 VTD-----VIADRPB--PVIRQGPNQNIVA VD--GTFFVLSCVATGSPVPTIL 487

QY   1430 WFHGGQPIVTTATGLTHIIAAGQILOVANLSGGSGOFECSCLAONEAGVLMQMKNASLIQDY 1489
     :|:| | | | | | | | | | | | | | | | | | | | | | |
Db    488 WRKDGVLTGSTQDSRIKQL--ENCVLQIRAKLGDTRYTCIASTPTSPEATWSAYIEVQRF 545

QY   1490 WWSVDRLATCASCCNRGVGOQPKRLCLNSTEYNPAHCACKVRPAQPIA----- 1539
     ||| | | | | | | | | | | | | | | | | | | | |
Db    546 -----GVPVQPPER-----PTDENLPISAPS-KPEVTVSRNVTIVLSWOIP 583

QY   1540 -CNRRCDCPMWMVTSWSACT-----RSCGGGVOTRRVTCTOKKL-----ASGIS 1581
     | | | :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

```

504 QY KLPWFQAOERLEBGAIVSEEPFIPANSACTVTCGVTOVRIVRCOVLLSFSQSVDLP 563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
501 Db --DWI-----TWSPVSCGMGRSR-----ERYVKQFP 527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
564 QY ID--ECGPKPASORACYAGPCSGE---LPEENPDETDGLFGLODFDELXDWRYEGFTK 618
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
528 Db EDGSMCKVPTTEKTCIVNEECSPSSCLVTEWG-----EWD-----E 564
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 QY CSBSCGGGVOEAVVSCINKOTREPAENLC--VTSRRPPQLLKSNLDPC--PARWEIGKW 675
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
565 Db CSASCGTGMRK--HRMIKW--PADGSMCKAETTEAEKMMPECHTIPCLLSPW--SEW 618
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
676 QY SPCSLLTCGVGLQTRDFVCSHLLSRENNETVILADELCRQPKPSTVQACNRNCP--P 730
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 Db SDCSVTCGKGNRTR-----ORMLKSAELGD--CNE-ELEQAEKMLPECPIDCELT 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
731 QY AWYTAQWQPSRFTCGGVOKREVILKORMADGSLFELPTFF-----C----- 772
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
668 Db EW--SOWSECNTSCGKGHMLTRFMKIEPQFGG-TACPETVQTRKCRVKLCRPGCKMRK 724
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
773 QY ---SASKPACQOACKDD-----CP-SWLLSDMTECSTSCGGETQTRSAICRKMILT 821
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
725 Db RWEKAREKRRSEQAKNLDNEQYVPCRKPW--TAWTECSTLCGGGIQERYMMVKRRKS 782
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
822 QY GLSTVVNWTLCPPILPFSSSRHPCMATC 849
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
783 Db -----TOFTSKD---KKELRACNVHPC 802
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

14
US-08-313-288B-12
; Sequence 12, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Duham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-313-288B-12

```

### Query Match

2.98; Score 277.5; DB 1; Length 802;

```

RESULT 15
PCT-US93-03164-12
; Sequence 12, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
;

```

Search completed: July 24, 2002, 04:18:21  
Job time: 6809 sec





Result No.	Score	Query %		Length	DB	ID	Description
		Match	Length				
1	1145.5	11.9	2165	2	T21371	hypothetical prote	
2	1088.5	11.4	1059	2	T22545	hypothetical prote	
3	887	9.3	1558	2	C89114	protein C37C3.6a [	
4	887	9.3	2167	2	T34395	hypothetical prote	
5	732.5	7.6	951	2	T00260	hypothetical prote	
6	639.5	6.7	1205	2	T18517	procollagen N-endo	
7	597	6.2	951	2	T00017	gene ADAMTS-1 prot	
8	584.5	6.1	550	2	T47158	hypothetical prote	
9	566.5	5.9	898	2	T14764	hypothetical prote	
10	468.5	4.9	1444	2	T18856	angiogenesis inhib	
11	422	4.4	837	2	T00335	hypothetical prote	
12	373	3.9	788	2	T25061	hypothetical prote	
13	372	3.9	4391	2	A38096	perlecan precursor	
14	352.5	3.7	5175	2	T20992	hypothetical prote	
15	352.5	3.7	5198	2	T43290	hemictenin precurs	
16	341	3.6	957	2	T15976	hypothetical prote	
17	332.5	3.5	1906	1	S68235	myosin-light-chain	
18	326	3.4	3707	2	S18252	heparan sulfate pr	
19	320.5	3.3	1259	2	S36136	neural cell adhesi	
20	318.5	3.3	7962	2	T38466	elastic titin - hu	
21	317.5	3.3	6642	2	T29757	protein UNC-89 - C	
22	317	3.3	1584	2	T00026	brain-specific ang	
23	312.5	3.3	1572	2	T00027	brain-specific ang	
24	311.5	3.2	469	1	S29126	properdin precursor	
25	311.5	3.2	1447	2	A54100	connectin/titin -	
26	309.5	3.2	4162	2	T42633	tumor suppressor p	
27	307.5	3.2	1260	1	S05479	neural cell adhesi	
28	302.5	3.2	1239	1	A32579	neuroglian - fruit	
29	301	3.1	1427	2	I51669	tumor suppressor -	



Db 104 --ANFLWRVDDGTPCOAATSAVCSKSGQIVGCDGLISSFRFDACGVCGGRGDTIC--- 158  
QY 189 RGOYKQSLSATKSDDDTVVAIPGYSRIIRLVKGPDLHYLETKTLOGTKGNSLSSTGIFL 248  
Db 159 ----- 158  
QY 249 VNSSVDFOKFPDKIELRMAGPLTADFIKIRNKGSSADSTVQFIFQYPIIHRKRETFDPP 308  
Db 159 -DNKG-----FIWKV----- 308  
QY 309 CSATC-----GGYOLTSAE-----CYDLRSNRVADQYCHYTPENIKPKPKQECNL 356  
Db 174 CASNDDDIVDWSGAGRSTASTQPIVVCVNAITGRVPEKLC---ADKLPRKVEARPCPM 230  
QY 357 DCPASDGYKQIMPYDLYHPLPRWEATPWTACSSCGGIGQISRAVSCVE--EDIQGHVTS 414  
Db 231 LICPS-----RWAAOWTECVPHCGEGTRKREYICVQTAHNVTVH--- 270  
QY 415 VEHWKCYTPMPKIAOPCNIFDCPKWLAQEMSPCVTTCGQGLRYRVVLCIDHRGMHTGG- 473  
Db 271 VPDTCFENGTRPAABENCVSTSCGWEAGKSKCTASCQGVRRHVACV-----GG 323  
QY 474 -CSPKTRHKEECIVPPCPYKPEKLEVEAKLPWFOAQELEGAAVS-----EPPS 525  
Db 324 CDDEGGRPRQETTCYAGIPCSIATN-----SLDNDRAY-LOGNTFGSDNHNNDWQAPR 376  
QY 526 PIPEAWSACTVTCGVGTQVIRVRCQVLLSFSQSADLPIDCEG-PKPASORACYAGPCPS 584  
Db 377 LVAGEWSTCSSTCGTGYMSRIVECVANPISAPIKPMSECDQEQKLFESCEVRS- 435  
QY 585 GEIPEFNPDETDLFGGLQDFDELYDMYEYEGFTKCSGSGGGOEAVVCSLKNKOTREPAP 644  
Db 436 -----PQED-----SKLSEDAEYQWRYGDTQCSASCLGGKKAALKCIQVSTGKSVQ 485  
QY 645 ENLCVTSRRPOLLKSNLPCPARWELGKWSQSLTCGVGLQTRDYFCSHLLSRE--MN 702  
Db 486 WSQCDARRRPPKSRPCNQHPCPFWITSKYSDCMSGSGTARRSVKCAQTGSKTDGAD 545  
QY 703 ETVLADELCRQPKPTQVACNRFCNPAMY-----PAWQPCSRCTGG 746  
Db 546 AHIVLRDRCHERKPKQETETCNVACPATWVSLNKRHNKIKLNKLTAKTQWTECSRSDS 605  
QY 747 GYOKREVLCORMADGSELELPTFCSA-SKPACQOACKKDDCPSEWLLSDWTECSTSCG 805  
Db 606 GERROVQWCEIRDSRGKTORRDPVECDANKP----- 637  
QY 806 EGTQTRSAICRMKLTGLSTVYNSTLCPLPFSSSRPCMLATCARGRPSTKHSPIIA 865  
Db 638 -----QIVEVCSFGCSRPE-----LL 654  
QY 866 ARKVIQTRRQKLFHVGGFAYLLPKTAVLRCPARRRKPLITWEKGOHLISSTHVT 925  
Db 655 SNRVFEQNAEQKLTGIGGVATLQGTGSIKCPAKKFTDKKIYKWKNGKIKNDAAHK 714  
QY 926 VAPFGVLTHLRKPDSAGVYTCAGPAREHFKVLIGGNRKLVARPLSPRSEEVLAGRK 985  
Db 715 VSANGLNRFVHARMEDAGYEC----- 736  
QY 986 GGPKEALQTHKHQNGIFSNKAEKGLAANPGSRYYDDLVSRLLEGGWPGELLASWEAQ 1045  
Db 737 ----- 736  
QY 1046 DSAERNITSEEDPGAEQVLLHLPFTMTVEQRRLDILGNLSQOPELRLDLYSKHLVAQLA 1105  
Db 737 -----FI-----DRLOQNVIT----- 746  
QY 1106 QEIFRSHLEHODTLKPSERTSPVTLSPHKHVSFGFSSSLRTSSTGADGSGRRRPHRPT 1165  
Db 747 -----LNFYRDF-----PASR----- 758  
QY 1166 ILRKISAAQLSASVTVHLGQTVLASGTLVLLHCEAIGHPRPTISWARGNEVEQFSD 1225  
Db 759 -----VOLA-----PKPQIPSTNQRQVQSK 780

QY 1226 RILLOPDDSLQILAPVADVGYFTCNATNALGYDSVSIATVLAGKPLVKTRMTVINTEK 1385  
Db 781 EDVIREQASV----- 790  
QY 1286 PAVTVDIGSTIKTVQGVNVNTINCOVAGVPAEAVTWFRNKSGLGSPHHLHBSLLLTNVSS 1345  
Db 791 -----LHK-----MNVSL 798  
QY 1346 SDGLYSCRAANLHGETESTQLLIDPPQVPTQLEDIRALLAATGPNLPSVLTSLGTQ 1405  
Db 799 -----IEALL--TAPN----- 807  
QY 1406 LVLDPCNSALLOCPKIGHPVNPITWPHGGQPIVATGLTHHTLAAQOILQVANLSGSG 1465  
Db 808 ----- 807  
QY 1466 EFSCLAQNEAGVLMOKASLVIOYMWSDRLATCSACNRCGVQOQPLRCLLNSTEVNPA 1525  
Db 808 -----DEKAREQLRY-----GNEIV----- 823  
QY 1526 HCAGKVRFAVQPIACNRDCPSRMMVTSWAS-TRSCG-GGVQTRVTCQKILKASISTP 1583  
Db 824 -----ARWDIGHWSECROKTVHAGVQARGISC-KVTFHGEIRN 861  
QY 1584 VSNDMCTOVAK-RPVDTQACNQOLCVERAFSSWGOCNGP-CIGPHLAVQHROVFCQTRGD 1641  
Db 862 VDSICESIASVRPPTRETPCHREDCPRWEASQWSECSQRCVSSMLAOKRRNVTCRTFNG 921  
QY 1642 ITLPSEQSALPRPVSTONCWSEAGSVHWRVSLWTLCTATCGNYGFSQRRRVECVHARTNK 1701  
Db 922 TSDVQIHCIDIIRPATIMDCPNQCKAEWRTSDWGSCESEGTGCVGLRLLSVWISSGR 981  
QY 1702 AVPEHLCSWGRPPANWQRC-----NITPC-----ENMECRDTTRYCEKVQKQLKQL 1748  
Db 982 PAGRN-CEQMRPHSARACVADEPLPCMPPTASALYQDASCQDSQSFCDIILFHSCDS 1040  
QY 1749 SDFKSRCCGTC 1759  
Db 1041 LEVRQKCCSTC 1051

## RESULT 3

C89114

protein C37C3.6a [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001

R:Accession: C89114

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: C89114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1558 &lt;STO&gt;

A:Cross-references: GB:chr\_V; PIDN:AAC25867.1; PID:g3294501; GSPDB:GN00023; CESP:C37C

C:Genetics:

A:Gene: C37C3.6a

A:Map position: 5

## Query Match

Best Local Similarity 9.3%; Score 887; DB 2; Length 1558;

Matches 229; Conservative 107; Mismatches 274; Indels 178; Gaps 29;

QY 27 RSEEDRDGLDAGWPWSECSRTCGGASYSRLRCLSSKSKSECRNIRYRTCSNVDCPPPEAG 86  
Db 70 KSGQKETGNWGPWVPENECSRSGGVQLEKRCQ--SGDCTGASVRYISCNLNAC--ESG 125  
QY 87 -DFRAQCCSAHNDVKHGHQFYEWLFPVSNPDNPCLKQAKGTTLLVLELAPKVLDRICY 145

A:Experimental source: strain Bristol N2; clone C37C3

C:Genetics:

A:Gene: CESP:C37C3.6b; CESP:C37C3.6a

A:Map position: 5

A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 15

Db 126 TDFRAEQCKFNDALDGNHKKPYKG--KNKCELVCKPESGNFYKWKADKVVDTGKCD 183  
QY 146 TESLDMCISGLCOIVGCDHQLGTVKEDNGVCGNGDGTCTRLVRGOY-KSOLSATKSDDT 204  
Db 184 SKSNDICVDGECPLVGCDDGLSSLFKDKCKDGDGCTKTIEGRFDERNLSPGYHD-- 241  
QY 205 VVAIPYGSRIIR-----LVLK-GPDHLYLETKLQGTGKGENSLSTGTFLVDNSS 253  
Db 242 IIKLPEGATNIKIQPARKSTNNLALKNGBDHFYLNGLIQLVEKEVEVGTT-IFVYDDA- 299  
QY 254 VDFQFPDKEILRMAGPLTADIVKIR-NSGSADSTVQFIFYQP-----IIHRWETDFEP 308  
Db 300 -----EPETLSAOGPLSEELTVALFRKGSRDATKYEFSIPLEEEVDYMYKFDNWT 352  
QY 309 CSATCGGGYQLTSACCYDLRSNRVADQYCHYYPENIKPKPKLOE-CNLDPCEASDGYKQ 367  
Db 353 CSVSCGKGVQTRNLKICIDGNKGRVEDDLCE-----ENNAIKPEFEKSCETVDC 403  
QY 368 IMPYDLYHPLPWEATPWTACSSCGG-GIOSRAVSCVEEDIQGHVTSVEENKCMYTPKM 426  
Db 404 -----EWFTGDWESCSSTCGDQGOQYRVYCHQVFANGRRVTVEDGNCCT-VERP 451  
QY 427 PIAOPCNIFDCPKWLAQEWSPCTVTCGGGLRYRVVLCIDHRGMHTGCGSPKTKPHKEEC 486  
Db 452 PVKQTCNRFACPEWQAGPWSACSEKCGDAFOYRSVTC-----AC 488  
QY 487 IVPTPCYKPKELPVEAKLPWFQKQAELEGAASVEEPSFIPAWNSACTVTCGVGTQVRI 546  
Db 489 -----RSEKEGEGKLLAAD-----AC 505  
QY 547 VRCQVLLSFQSVADLPIDCEGPKPASORACYAGPCSGEIPFNPDETGLGGLQDFD 605  
Db 506 -----PADEQE--KFTDERTCNLGPCEGL----- 527  
QY 607 ELYDWEYEGFTKCSSECGGVQAEVAVVSCLNKQTRP-PAENLCVTSRRPPQLLKSC-NLD 664  
Db 528 TFFTGEWNLCTRCNDT-----EETREVTCKDSQGRAYPLEKCLVDNSTEIPTRSCATQP 583  
QY 665 PCPARWEIGKWSPCSLTCGVLQTRDVFCSHLLSREMNETVILADE-LCROPKPESTVQAC 723  
Db 584 PCEYEWTVSEWSKCTTECGHGHKTRVIC-----AIHONGGLEVDDEGHCHQAEKPEGKTNC 639  
QY 724 -NRFNCPAWYPAQWQPCSRCTCGGGVQKREVLCQRMADGSFLELPETFCSSASKPACQQA 782  
Db 640 TNEEKTCTGTWTSWSECTAECGGGSDQDRVAVC-----LNYDKKVPPEWCD EAVKPEKQD 695  
QY 783 CKKDDCPS 790  
Db 696 CNVDDCPT 703

RESULT 4  
T34395  
hypotheical protein C37C3.6b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
R:Accession: T34395; T34394  
R:Geisel, C.; Bradshaw, H.  
submitted to the EMBL Data Library, July 1996  
A:Description: The sequence of C. elegans cosmid C37C3.  
A:Reference number: Z21518  
A:Accession: T34395  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2167 <GEI>  
A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b  
A:Experimental source: strain Bristol N2; clone C37C3  
A:Accession: T34394  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1555, /SKF' <GE2>  
A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a

Query Match 9.3%; Score 887; DB 2; Length 2167;

Best Local Similarity 29.1%; Pred. No. 1.4e-42;

Matches 229; Conservative 107; Mismatches 274; Indels 178; Gaps 29;

QY 27 RSEEDRDGLWDAGWPMSECSRTCCGGGASYSRLRCLSKSCGEGNIRYRCSNVVDCPPEAG 86

Db 70 KSKQETGNGWPPVNECSRCGGGVQLEKQCC--SGDCTGASVRYISCNLAC--ESG 125

QY 87 -DFAQQCSAHNDVHHGQFYEWLPVSNDDPDCSLKCOAKGTTLLVELAPKVLDTGRCY 145

Db 126 TDFRAEQCKFNDALDGNHKKPYKG--KNKCELVCKPESGNFYKWKADKVVDTGKCD 183

QY 146 TESLDMCISGLCOIVGCDHQLGTVKEDNGVCGNGDGTCTRLVRGOY-KSOLSATKSDDT 204

Db 184 SKSNDICVDGECPLVGCDDGLSSLFKDKCKDGDGCTKTIEGRFDERNLSPGYHD-- 241

QY 205 VVAIPYGSRIIR-----LVLK-GPDHLYLETKLQGTGKGENSLSTGTFLVDNSS 253

Db 242 IIKLPEGATNIKIQPARKSTNNLALKNGBDHFYLNGLIQLVEKEVEVGTT-IFVYDDA- 299

QY 254 VDFQFPDKEILRMAGPLTADIVKIR-NSGSADSTVQFIFYQP-----IIHRWETDFEP 308

Db 300 -----EPETLSAOGPLSEELTVALFRKGSRDATKYEFSIPLEEEVDYMYKFDNWT 352

QY 309 CSATCGGGYQLTSACCYDLRSNRVADQYCHYYPENIKPKPKLOE-CNLDPCEASDGYKQ 367

Db 353 CSVSCGKGVQTRNLKICIDGNKGRVEDDLCE-----ENNAIKPEFEKSCETVDC 403

QY 368 IMPYDLYHPLPWEATPWTACSSCGG-GIOSRAVSCVEEDIQGHVTSVEENKCMYTPKM 426

Db 404 -----EWFTGDWESCSSTCGDQGOQYRVYCHQVFANGRRVTVEDGNCCT-VERP 451

QY 427 PIAOPCNIFDCPKWLAQEWSPCTVTCGGGLRYRVVLCIDHRGMHTGCGSPKTKPHKEEC 486

Db 452 PVKQTCNRFACPEWQAGPWSACSEKCGDAFOYRSVTC-----AC 488

QY 487 IVPTPCYKPKELPVEAKLPWFQKQAELEGAASVEEPSFIPAWNSACTVTCGVGTQVRI 546

Db 489 -----RSEKEGEGKLLAAD-----AC 505

QY 547 VRCQVLLSFQSVADLPIDCEGPKPASORACYAGPCSGEIPFNPDETGLGGLQDFD 606

Db 506 -----PADEQE--KFTDERTCNLGPCEGL----- 527

QY 607 ELYDWEYEGFTKCSSECGGVQAEVAVVSCLNKQTRP-PAENLCVTSRRPPQLLKSC-NLD 664

Db 528 TFFTGEWNLCTRCNDT-----EETREVTCKDSQGRAYPLEKCLVDNSTEIPTRSCATQP 583

QY 665 PCPARWEIGKWSPCSLTCGVLQTRDVFCSHLLSREMNETVILADE-LCROPKPESTVQAC 723

Db 584 PCEYEWTVSEWSKCTTECGHGHKTRVIC-----AIHONGGLEVDDEGHCHQAEKPEGKTNC 639

QY 724 -NRFNCPAWYPAQWQPCSRCTCGGGVQKREVLCQRMADGSFLELPETFCSSASKPACQQA 782

Db 640 TNEEKTCTGTWTSWSECTAECGGGSDQDRVAVC-----LNYDKKVPPEWCD EAVKPEKQD 695

QY 783 CKKDDCPS 790

Db 696 CNVDDCPT 703

RESULT 5

T00260

hypotheical protein KIAA0605 - human

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000

Query Match	6.7%;	Score 639.5;	DB 2;	Length 1205;
Best Local Similarity	22.8%;	pred. No. 1.1e-28;		
Matches 208;	Conservative 96;	Mismatches 264;	Indels 345;	Gaps 34;
QY	32	RDGLDWAQFMSRCRSCGGSYSLLRC-----LSSKSCGRNIIRYTCNSNDCPPEA	85	
Db		: : :       : : :     : : :		
QY	554	ROGNWGAWSFGSCRCGTCGGVGFERTQOCNPHANGRTCSGLAYDFQLCNSQDCPDAL	613	
Db		: : :       : : :     : : :		
QY	86	GDPRAOCS-----AHNDYKHKGFEYEWLPVSN-DPDNFCSLKCOAKGTTLVYELAPKV	138	
Db		: : :       : : :     : : :		
QY	614	ADFREQCQRQWLDYFEHGDQAQHH-----WLPHEHROAKERCHLYCESKETGEVVSMAKRV	668	
Db		: : :       : : :     : : :		
QY	139	LDGTRC-YFESLDMCLISGLCQIVCCDHQIGLSTVKEDNCVCGNDGSTCRIVRQYK---S	194	
Db		: : :       : : :     : : :		
QY	669	HGTRCSYKDAFSLCVRGDCRKGCCDGVITGSSQDKCGVCGGDNSHCKVKVGTFSRSPK	728	
Db		: : :       : : :     : : :		
QY	195	QLSATKSDTTVAIPVGSRHIRLVLKGPDHLXYLETTLQGTK-----GENSLS-STGTFV	249	
Db		: : :       : : :     : : :		
QY	729	KUGYIK---MFEIPAGARHLIIQEQADTTSHLAVKNLEIGKFIENENDVDPNSTKTFIA	784	
Db		: : :       : : :     : : :		
QY	250	DNSSVDVDFKFPDKEILRMAGPLTAQFIVKIRNSGSDSTVQFIYQPIIHRMRETDFFPC	309	
Db		: : :       : : :     : : :		
QY	785	MGVWEYRDEQRETLQTMGPHLGTITVLVIPEGDARTS---LTKYMIHE-----	832	
Db		: : :       : : :     : : :		
QY	310	SATCGGGQLTSAECYDILRSNRVADQYCHYYPENIKPKPLQECNLDPCPASDGYKQIM	369	
Db		: : :       : : :     : : :		
QY	833	-----DSLNVDDNNVLEDD-----SVGYE-----	851	
Db		: : :       : : :     : : :		
QY	370	PYDLYHPLPRWEATPWTACSSCGGIGQSRVSCVEEDIQGHVTSVEEWKCMYTPKMPIA	429	
Db		: : :       : : :     : : :		
QY	852	-----WALKWSPSCFPCGGSGQFTKYKC-----	875	
Db		: : :       : : :     : : :		

QY 430 QPCNIFCPKWLAEWSPCTVTCGGRLRYRVVLICIDHRGHTGSCPKTKPH-IKEECIV 488  
Db 876 -----RRR-----LDHKMVRHGFCDSVSKPAIRRTG-- 902  
QY 489 PTPCYKPKELPVEAKLPWFQKQAELEGAASVEEPSFIPRAWSACTVTCG-VGFOVRIV 547  
Db 903 -----NPOE-----CSQPVVWTGEWEPSCRSRSGRTGQVRVS 934  
QY 548 RCQVLLSFSQSADLPIDCEGPKPASORACYAGPCSGEIPFNFDEIDGLFGGLQDFDE 607  
Db 935 RC----- 936  
QY 608 LYDWEYEGTKESCGGQVQAEVYVCLNKQTRPAEENLCVTSRRPPOLLKSCMLDPCP 667  
Db 937 -----VOPLHNNTTRSVHYTKHNDAR--PGRACNRELCP 970  
QY 668 ARWEIGKWSPCSLTCGVGLQTRDVFCSHLLSREMNETVILADE---LCRQPKSPVQACN 724  
Db 971 GWRAGSWSQSVTCGNGTQERPVLCR-----TADDSFGVCREERPETARICR 1018  
QY 725 RFNCP-----PAWYPAQW-----OPCSRT-----CGGGVOKREVLCCKORMADGSFL 765  
Db 1019 LGPCPRNTSDPSKSKSVVQWLSRPDPNSPVQETSSKRCQG---DKSVFCR-----M 1067  
QY 766 ELPEPFCASAKPACQACQCKDCPSEWLLSDWTECSTGCGEGTQTRSAICRMLKTLGLST 825  
Db 1068 EVLSRYCSI--PGYNKLCCKSCNPHNL-----TDVDRAPPPSGKHNDIEELMPTLSVPT 1121  
QY 826 VNSTILCPP-----LPFSSIRCMATCARPKPSTK-----HSPHIAA 865  
Db 1122 LYMEVQPPGPIPLEVPLNTS-----STNATEDHPETNAVDVPYKIPGLEDEVPNNLIP 1175  
QY 866 AR-KVYIQROR 877  
Db 1176 RRPSPYEKTRNQ 1188  
RESULT 7  
gene ADAMTS-1 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
A:Accession: T00017  
R:Kuno, K.; Lizaasa, H.; Ohno, S.; Matsushima, K.  
Genomics 46, 466-471, 1997  
A:Title: the exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene  
A:Reference number: Z14055; MUID:98110583  
A:Accession: T00017  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-951 <KUN>  
A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057  
A:Experimental source: strain 129SVJ  
C:Genetics:  
A:Gene: ADAMTS-1  
A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2  
C:Superfamily: thrombospondin type 1 repeat homology  
F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>  
Query Match 6.2%; Score 597; DB 2; Length 951;  
Best Local Similarity 29.1%; Pred. No. 2.2e-26;  
Matches 143; Conservative 51; Mismatches 180; Indels 118; Gaps 16;  
QY 34 GLWDAGWPWSECSRTCCGGGASYSLRRC-----LSSKSCGGRNIRYR 86  
Db 544 GSWGPWGPWGDCSRTCCGGGVQVYTMRECDNPVPKNGKYGCEGRVYRSCNIEDCPDNNK 603  
QY 87 DPAQCCSAHNDVKH----HGQFYEWLP--VSNDDPNPCSLKCOAKGTTLVVELAPKVID 140  
Db 604 TFEEOCEAHNEFSKASFGNEPTEVTPKYAGVSPKDRCKLICEAKIGYFFVLOPKVD 663  
QY 141 GTRCYTESLDMCISGLCQIVGCDHGLGTVKEDNCVCNGDGTCTCLRVGRQYKSQLSATK 200

Db 654 GTPCSPDSTVCCVQGCQVKAGCDRIIDSKKFDKCGVGGNGSTCKKMSG---IVTSTRP 720  
QY 201 SDDTVVAIPYGRHTR-----LVLGKPDHLYLETWLOQTKCENSISSTG 245  
Db 721 GYHDIIVTIPAGATNIEVKHRNQSRNNGSFLAIRAADGTIYINGNFTLSTLEQDLTYKG 780  
QY 246 TFL-VDNSSVDQKFPDKELLRMAGPLTADFIVKIRNSGSA-DSTVOFIYQYPIIHRWE 303  
Db 781 TVLIRSGSSAALEIR-----IRSFPLKEPLTIOVLVGHALRPKIKFIYFM----- 826  
QY 304 TDFPFCATCGGYQLTSAECYDLRSNRVADQYCHYYPENIKPKPKLQECNLDPCPASD 363  
Db 827 -----KKKTESFNAIP----- 837  
QY 364 GYKQIMPYDLIPLPWEATPWTACSSCGGTQSRVSCVEEDIOGHVTSVEWKMCT 423  
Db 838 -----TFSEWITEWEGECSKTCGSGWQRVVQC--RDINGHPAS-----ECAKE 879  
QY 424 PKMPIAOPCNIFDCPKWLAEWSPCTVTCGGRLRYRVVLICIDHRG--MHTGGCSPKTKP- 480  
Db 880 VKPASTRPACDLPCPHWQGDWSPCKTCCKGYKKRTLKCVSHDGGVLSNESCDPLKKPK 939  
QY 481 HIKEECIVPTPC 492  
Db 940 HYIDFCTL-TQC 950  
RESULT 8  
T47158  
hypothetical protein DKFZp762C1110.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
A:Accession: T47158  
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: 224379  
A:Accession: T47158  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-550 <AAA>  
A:Cross-references: EMBL:AL162080  
A:Experimental source: adult melanoma (Mewo cell line); clone DKFZp762C1110  
C:Genetics:  
A:Note: DKFZp762C1110.1  
Query Match 6.1%; Score 584.5; DB 2; Length 550;  
Best Local Similarity 29.4%; Pred. No. 5.6e-26;  
Matches 148; Conservative 59; Mismatches 168; Indels 129; Gaps 19;  
QY 23 SRTARSEEDR--DGLWDAGWPWSECSRTCCGGGASYSLRRC-----LSSKSCGGRNIRYR 74  
Db 130 NKTDRKHEDTPPHGSGWMPWGDCSRTCCGGGVQVYTMRECDNPVPKNGKYGCEGRVYR 189  
QY 75 TCSNVDCPEAG--DFAQOCCSAHNDVKH----HGQFYEWLP--VSNDDPNPCSLKCOAKG 127  
Db 190 SCNLEDCPDNNKCTFEEOCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICEAKG 249  
QY 128 TTLVVELAPKVIDGTRCYTESLDMCISGLCQIVGCDHGLGTVKEDNCVCNGDGTCTCL 187  
Db 250 IGYFFVLOPKVDGTPCSPDSTVSCVQGCVRKAGCDRIIDSKKFDKCGVGGNGSTCKK 309  
QY 188 VRGQYKSQLSATKSDDTVVAIPYGRHIRLVKGPDLHLYETWLOQTKCENSISSTG 243  
Db 310 ISG---SVTSAPGYHDIITPTGATNIE-----VKQRNGSRNNGSFLAIKAA 356  
QY 244 TGTFLVDNSSVDQKFPDKELLRMAGPLTADFIVKIRNSGSA-DSTVOFIYQYPIIHRWR 302  
Db 357 DGTYY-----LNGDYTL-----STLEQDIIYKGVVLR-- 383  
QY 303 ETDFPFCATCGGYQLTSAECYDLRSNRVAD-----QYCHYYPENI 345

Db 384 -----YSGSALERIRSESPKLEPLTIOVLTVGNALRPKIKYIFV----- 425  
QY 346 KPKPKLOECNLDPCPADGKYQIMPYDLYHPLPRWEATPWTACSSCGGGIQSRVSCVE 405  
Db 426 --KKKKESENAIP-----TFSAWVIEEWGECSCGELGWQRLLVBC-- 464  
QY 406 EDIQGHVTSVEBWKMYTPKMPIAQPCNIFDCPKWLAQEWSPCTVTCGGGLRVKRVLCID 465  
Db 465 RLINGOPAS-----EKAKEVKFASTRPCADHPCQWOLGSEWSSCKTCKGKXKSLKLS 520  
QY 466 HRG--MHTGGCSPKTKP--HIKEEC 486  
Db 521 HDGGVLSHESCDPLKKPKHFIDFC 544

RESULT 9  
T14764  
hypotheical protein DKF2p434H204.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14764  
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: 218181  
A:Accession: T14764  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-898 <WAM>  
A:Cross-references: EMBL:AL110226  
A:Experimental source: adult testis; clone DKF2p434H204  
C:Genetics:  
A:Note: DKF2p434H204.1

Query Match 5.9%; Score 566.5; DB 2; Length 898;  
Best Local Similarity 23.3%; Pred. No. 1.1e-24;  
Matches 279; Conservative 118; Mismatches 404; Indels 397; Gaps 63;

QY 608 LYDWEYEGFTKSESCGGGVQAVVSCNLKQTRPAENCLVTSRRPQLLKSNLDPCP 667  
Db 34 VFSWHYGPWKIVCTVTCGRGVQNVYCLERQA-GPVDEEHCDPLGRPDQQRKCEQPCP 92  
QY 668 ARNEICKWSPCSLTGCV-GLOTRDFVCSHLLSREMNETVILADELCRO-PKPSTVOACNR 725  
Db 93 ARWAGGEWQJCSGCGGLSRRAVLG--IRSVGLDEQSALEPPACEHLPRPTETPCNR 150  
QY 726 -FNCPPAWYPAQWQPCSRCTGGGVQKREVLCKORMADGSFLELPETFC-SASKPACQAC 783  
Db 151 HVPCTPATWAGVNNSSQSVTCGEGTQRNVLCNTDTC-----VP---CDEAQQPASEVTC 201  
QY 784 KKDCPSEWLLSDWTECSTSCGEGTQTRSAICRKMVLKTLGLSTVWVNSTLCP-PLPSSSIR 842  
Db 202 SLPLC--RWPLG--TLGPEGSGSSSHELF-----NEADFIPHLAPRSP-ASSPK 249  
QY 843 PCMIATCARCPRPSTKXSPHIAAARKVYIOTRQRKLFHVVGFPAYLLPKTAVLURCPAR 902  
Db 250 PGTW-----GNAIEEAEPLDLPGVF-----VDDFY----- 277  
QY 903 RVKRLITWEKQHLISSHTVIVAPGYLKTIRLKPSSDAGYVTCAGPAREHVIKLI 962  
Db 278 -----DYNFINPHE-----DLSYGUSEEPDLDLA 303  
QY 963 GNRKLV--ARPLSPRSEEVLAGKGPKEALQTHKHQNGIFNSGKA---EKRLAAMP 1017  
Db 304 GDRTPPHSRPAAPSTGSPVATEPPAAKE-----EGVLGPWSPSPWSPQAGRSPPP 355  
QY 1018 GSRVYDLVSLRLEGGWPGELLASWEAQDSARNNTTSEEDPCAEQVLLHLPF----- 1069  
Db 356 PS-----EQ--TPGNFLI-----NFTPEETPGADPLGLPSLWPRVST 393  
QY 1070 ----TWYEQRRLLDILG--NLQQPEELRDLYSKHLVAQLAQIFRSHLEHQDTLLKPS 1123  
Db 394 DGLQTPATPESQNDFFVPGKDSQSLPPPPWRD-----RTNEVFKDDEPK----- 437

QY 1124 ERTSPVTLSPKHVSGFSSSLRTSSGCDAGGSRPHKPTILRKLSAAQOLSASEVVT 1183  
Db 438 -----GRGAPHLPP-----RFSSTLPPLPSP----- 457  
QY 1184 HLGOTVALASGTLVLLHCEAIGHPRPTISW--ARNGEEVQFSDRILLQPDOSLQILAPV 1241  
Db 458 -VGSTHSSPSPDVAEL-----WTGCTVAMEPALEGG-----LGPV 491  
QY 1242 EADVGFYTCNATNALGVDYSIAVTLAGKPLVKTSRMTVINTTEKPAVTVDI-----GS-T 1295  
Db 492 DSEL-----WPTGVGA-SLLPPPIAPLPEMKVDRDSSLERPTSPFPAPGSGWD 538  
QY 1296 IKTVQGVNVTINCOVAGVPEAEVTFWRNKSGLGSPHLLHGGSLLLNVSSDQGLYSCLA 1355  
Db 539 LQTV-----AVM-----GTFELPTLT-----GL----- 556  
QY 1356 ANHGEFTESTOLLILDPQVPTOLEDIRALLAANTGNPLSVLTSLPGTOLVLDPCNSAL 1415  
Db 557 -----GHMPEPA--LMPGKGPQESLS-----PEV--PLSSRLISTPAWDS- 593  
QY 1416 LGCPKIGHFVPVNTWFGGQPIVITATGLTHILAAGQILOVANLSSGSGQGEFSCLAQNEA 1475  
Db 594 ---PANSRVRPET-----QPLAFLSLA-----EA 613  
QY 1476 GVLMOKASILVIOYVWSVDRLATCSAGCNRGVQOPRLRLCLLNSTEVNPAHCAGKVRPAV 1535  
Db 614 G--PPADPLVVRNASQWAGNNSCSTTCGLGAVWRP-VRCSSGRDE---DCAPAGRP-- 664  
QY 1536 QPT-ACNRDCPSRMWVTSWASACTRSCGGVQTRRVTCQKLKASGIST-----PVSNDMCT 1590  
Db 665 OPARRCHLRPC-ATWHSNWSKSCSGSGSVRDVQC-----VDTRDLRLRPFHCQ 716  
QY 1591 QVAKRPEVDTOACNOQLCVWEAFSSWGQCGPCIGPHLAVQHVQECQTRDGTILPSEQCS 1650  
Db 717 PGAPKPAHRPCGAPQCLISWTISSHRESEACGG---GEQRLVTC-PEPGL-----CE 766  
QY 1651 ALPRPVSTONCSEACSVHVRVSLWTLTATCGNYGFSRRCVCHARTKAVPEH---L 1707  
Db 767 EALRPNTTTPCTHPTC-QWVVGPMGQCSAPCGG-GVORRLVKVNTQT--GLPEEDSDQ 822  
QY 1708 CSMGPPANQWQCNITPCENME---C---RDTTRYCEKVKQKLCQLSQFSKRCGCGTC 1759  
Db 823 CGHEAMPSSRRCGTCDEPVEPPRCERDLSFGCEITLRLGRCPLTIRTQCCRSC 880

RESULT 10  
T18856  
angiogenesis inhibitor homolog - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T18856; T24653  
R:McMurray, A.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: Z19031  
A:Accession: T18856  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1444 <WIL>  
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1  
A:Experimental source: clone C02B4  
R:McMurray, A.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: Z19917  
A:Accession: T24653  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1444 <W12>  
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1  
A:Experimental source: clone T07C5  
C:Genetics:  
A:Gene: CESP:C02B4.1  
A:Map position: X

A; Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 566		T00355		hypothetical protein KIAA0688 - human	
Query Match		4.4%;		Score 422; DB 2; Length 837;	
Best Local Similarity		22.0%;		Pred. No. 9.2e-19;	
Matches 207; Conservative 112; Mismatches 269; Indels 351; Gaps 54;		Accession: T00355		C; Species: Homo sapiens (man)	
		R; Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998		C; Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000	
22 SSRTARSEE--DRGLWDANGPSECSRTCGGASYSLSRCLSLKSCSEGRNIRYRTCSNV 79		A; Title: Prediction of the coding sequences of unidentified human genes. X. The compl			
774 SSKLQERPCDNGCWTNDEWSSCSQCGGRRYRIRKCLDDK-CDGDDLEKESCNTQ 832		A; Reference number: Z14142; MUID:98403880			
80 DCPPEA-GDFRAQCSAHNDYKHGQFYEWLPVSNPDNPKSLCKQAKGTTLVVELAPKY 138		A; Accession: T00355			
833 KCSQSGWDG-----WL-----PCSVSC-----GIGFOIRERL 859		A; Status: preliminary; translated from GB/EMBL/DBDJ			
139 LDGTRCYTESLDMCISGLCQIVGDHQLGISTVKED-----NCGVNGDG-----STCRILV 188		A; Residues: 1-837 <IS>			
860 CDGELCATANKO---ARTCNOQCPSATSLVSWSEGEWTTCSATCGEGLOSRERSCR-- 914		A; Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190			
189 RGQYKSLSATKSDTVVAIPYGRHRLVLKGP-DHLYL-----ETKTLQGTGKGENSLST 244		A; Experimental source: brain			
915 RG-----SCTEDD-----ASQTRCVNGPCEHSLYIWEWTTCTCSCSFDSRKRI 959		A; Gene: KIAA0688			
245 GTFLDSSVDFQKPKDEILMA-----GPLTADFIVKIRNSGSADSTVQFIYOP 296		C; Superfamily: thrombospondin type 1 repeat homology			
960 AK--CDGITENCQDKIDEETCDIACLRKHSGPISPR-----RPKLITSNDLRKAFGRP 1012		F; 519-575/Domain: thrombospondin type 1 repeat homology <THR3>			
297 I-----IHRWRETFPPCSATCGGQVQLTSAECYDLRNRVADQVCHYYPIKPKPL 351		Query Match		4.4%;	
1013 LLPIESIHSEKSEWGPCVTCGSGRRVTRGCOEAS-----C-----PEQHI 1055		Best Local Similarity		34.6%;	
352 Q--ECMLDPCASDGKQIMPLYLPLRWETATPWTACSSCG-GGIQSRAVSCVEEDI 408		Matches 100; Conservative 42; Mismatches 115; Indels 32; Gaps 11;			
1056 QTECNLNSC-----LELFIWSDSSCSKSCGQIGIQRKLCLEFNA 1098					
409 QGHVTSVEEKWMYTPKMPIAOPC-----NIFDCPKLA-QEWSPCVTCQGL 456					
1099 ECS-SVAESRCKDLFS-----CSSISSGRTISENGFDAPRWSESSWSACSFLTST 1151					
457 RYRVVICIDHRGHTGCGSPKTHIKKECIVPTPCYKPEKLVPKAEKLPWFOAQOELE 516					
1152 RRRECQVVD-----PTVQGC-----A 1168					
517 GAAYSEEP-----SFIPEA-----NSACTVTCG-VGTQVIRVRCOVLLSFSQSVADLP 563					
1169 GAIIEQIPCAPGSCPSAGGWSIWESSCSKCGDTHGQIR----- 1210					
564 IDECEGPKPASORACYAGPCSGEIPFNPDDETGLFGLODFDELYD--W-EYEGFTKCS 620					
1211 NRMCEPIP-SNRGAY---CSGSFDPQPCVMDNVC-----DEKVDGWTDTAWSECT 1261					
621 ESCGGVQVEAVVCLNKQTRPAENLCVTSRRPPQLLKSCLNDPC--PAR-----WE 671					
1262 DYCRNGHRSRTRFCANRPSGGQA--CTGS-----DELANPCFPAPCHLRDGGWS 1311					
672 I-GKWSPCSLTCGVGLQTRDFVCSHLLSKENNETVILADELCROPKPKSTVQACNRFNCP 730					
1312 TWSQMTWPCASCGFGVQTRDRSCS-----SPEPKGGQSCS----- 1346					
731 AWYPAQWPCSRTCGGGVQKREVLCKQRMADGSELELPETFCASKPACQACKKDDCPFS 790					
1347 -----GLAQHTSLC-----DLPAQDH-----ESDG 1366					
791 EW-LLSDWTECSTSGEGTQTSATCRKMLKTGLSTVNVNSTLCPLPFPSSIRPCW---- 845					
1367 ENSAWNEWSGCMGNCIGTRTRVAC-----VSPPV--SDGGQPCGFRSS 1409					
846 -LATCARCPKPTK-----HSPHIAAARKVYIQTREOR 877					
1410 EITECRQ--SPSTALCSSFTTSSHLADG--YFIDTDQOQ 1444					
RESULT 11					

hypothetical protein T21B6.3 - Caenorhabditis elegans  
C; Species: Caenorhabditis elegans  
C; Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C; Accession: T25061  
R; Cottage, A.  
submitted to the EMBL Data Library, November 1995  
A; Reference number: Z19975  
A; Accession: T25061  
A; Status: preliminary; translated from GB/EMBL/DBDJ  
A; Molecule type: DNA  
A; Residues: 1-788 <WIL>  
A; Cross-references: EMBL:268011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3  
A; Experimental source: clone T21B6  
C; Genetics:  
A; Gene: CESP:T21B6.3  
A; Map position: X  
A; Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2  
Query Match 3.9%; Score 373; DB 2; Length 788;  
Best Local Similarity 25.3%; Pred. No. 1.3e-13;  
Matches 133; Conservative 49; Mismatches 156; Indels 188; Gaps 29;  
341 YPENIKPKLQECNLDPCASDGYKQIMP---YDLVHPLPRWEATPWATCSSCGGGIQ 397



Db 429 YPTRYPAP-----PPPPACDGGCVNPPVSGVYVHD---W---SDWSTCSTCGDGAK 476  
 QY 398 SRVSCVEDIOGHVTSVEEWKMTKPKPIAOPCNIFDCPKWLAQ-EWSPCTVTTCGQGL 456  
 Db 477 SRRCSTNNCGADYETE-----PCNLGFCQWSEWSTCSASGSGQ 522  
 QY 457 RYRVVLICDHGMHTGGGSPKPKPKIKERCIVPTCYKPKKLPVAKLPWFKQAQELPEE 516  
 Db 523 RERTFRC-----HLGTRNCRGKDYEQCSA-GPC-----PEWSQWED--- 559  
 QY 517 GNAVSEEPSFTEPMSACTVTCGVGTQVRIVRQCVLSSFSQSVADLPIDE-CEGPKPASQ 575  
 Db 560 -----WGQCSVTCGGGVAVRQRFC-----LCGVFGDHLICQPKR-TEQ 595  
 QY 576 RACYAGPCSGRIPEFNPDDETGLGGLQDFELDYWEVETKSCSEGGGVQEAWVSC- 634  
 Db 596 RACDGGPCSLWSP-----WQENSTCSASCGSGMKRQRVCQ 631  
 QY 635 LNKOTREPAENLCVTSRRPQLKSCNLDPCPARW-EIGKWSPCSLTCGVGLQTPROVFC 693  
 Db 632 FGTDCCQGPNEES-----QFCYGPFC-AEWTEWCEWSGCSCSKGPGQRTTRGCG 678  
 QY 694 SHLLSRWNETVILADELCROPKPSTV-----QACNRFNCPWVPAWOPCSRCTCGGSGVQ 749  
 Db 679 LGPNQGEAT-----TCQGPSTETTLCEQSCCNWS---EW---CHWSMCDKCEGCG-- 723  
 QY 750 KREVLKORMADGSFLELPETFCASKPACQACKDDCPSEWLLSDWTEGST--SGEG 807  
 Db 724 -----QVRVIEYWF-----RTGC-----EWSPCSTQACEVG 750  
 QY 808 TQRTSATCRMLKTLGSTV-----VNSTLCPLPPESSIRPC 844  
 Db 751 VQSRQRC--VGESGCHICGLAESQCRGLTQCPKPAP-----PC 788

RESULT 13  
 A38096  
 perlecan precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 05-Nov-1999  
 C:Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736  
 R:Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.  
 J. Biol. Chem. 267, 8544-8557, 1992  
 A:Title: Primary structure of the human heparan sulfate proteoglycan from basement membrane, laminin, neural cell adhesion molecules, and epidermal growth factor.  
 A:Reference number: A38096; MUID:92233084  
 A:Accession: A38096  
 A:Molecule type: mRNA  
 A:Residues: 1-4391 <MUR>  
 A:Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427  
 R:Kallunki, P.; Tryggvason, K.  
 J. Cell Biol. 116, 559-571, 1992  
 A:Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD protein with cell adhesion molecules, and epidermal growth factor.  
 A:Reference number: A41736; MUID:92112994  
 A:Accession: S19256  
 A:Molecule type: mRNA  
 A:Residues: 1-57, 'D', 59-434, 'A', 436, 'FL', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908, 'R', 911-929, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-3634, 'R', 3635-3636, 'R', 3637-3638, 'R', 3639-3640, 'R', 3641-3642, 'R', 3643-3644, 'R', 3645-3646, 'R', 3647-3648, 'R', 3649-3650, 'R', 3651-3652, 'R', 3653-3654, 'R', 3655-3656, 'R', 3657-3658, 'R', 3659-3660, 'R', 3661-3662, 'R', 3663-3664, 'R', 3665-3666, 'R', 3667-3668, 'R', 3669-3670, 'R', 3671-3672, 'R', 3673-3674, 'R', 3675-3676, 'R', 3677-3678, 'R', 3679-3680, 'R', 3681-3682, 'R', 3683-3684, 'R', 3685-3686, 'R', 3687-3688, 'R', 3689-3690, 'R', 3691-3692, 'R', 3693-3694, 'R', 3695-3696, 'R', 3697-3698, 'R', 3699-3700, 'R', 3701-3702, 'R', 3703-3704, 'R', 3705-3706, 'R', 3707-3708, 'R', 3709-3710, 'R', 3711-3712, 'R', 3713-3714, 'R', 3715-3716, 'R', 3717-3718, 'R', 3719-3720, 'R', 3721-3722, 'R', 3723-3724, 'R', 3725-3726, 'R', 3727-3728, 'R', 3729-3730, 'R', 3731-3732, 'R', 3733-3734, 'R', 3735-3736, 'R', 3737-3738, 'R', 3739-3740, 'R', 3741-3742, 'R', 3743-3744, 'R', 3745-3746, 'R', 3747-3748, 'R', 3749-3750, 'R', 3751-3752, 'R', 3753-3754, 'R', 3755-3756, 'R', 3757-3758, 'R', 3759-3760, 'R', 3761-3762, 'R', 3763-3764, 'R', 3765-3766, 'R', 3767-3768, 'R', 3769-3770, 'R', 3771-3772, 'R', 3773-3774, 'R', 3775-3776, 'R', 3777-3778, 'R', 3779-3780, 'R', 3781-3782, 'R', 3783-3784, 'R', 3785-3786, 'R', 3787-3788, 'R', 3789-3790, 'R', 3791-3792, 'R', 3793-3794, 'R', 3795-3796, 'R', 3797-3798, 'R', 3799-3800, 'R', 3801-3802, 'R', 3803-3804, 'R', 3805-3806, 'R', 3807-3808, 'R', 3809-3810, 'R', 3811-3812, 'R', 3813-3814, 'R', 3815-3816, 'R', 3817-3818, 'R', 3819-3820, 'R', 3821-3822, 'R', 3823-3824, 'R', 3825-3826, 'R', 3827-3828, 'R', 3829-3830, 'R', 3831-3832, 'R', 3833-3834, 'R', 3835-3836, 'R', 3837-3838, 'R', 3839-3840, 'R', 3841-3842, 'R', 3843-3844, 'R', 3845-3846, 'R', 3847-3848, 'R', 3849-3850, 'R', 3851-3852, 'R', 3853-3854, 'R', 3855-3856, 'R', 3857-3858, 'R', 3859-3860, 'R', 3861-3862, 'R', 3863-3864, 'R', 3865-3866, 'R', 3867-3868, 'R', 3869-3870, 'R', 3871-3872, 'R', 3873-3874, 'R', 3875-3876, 'R', 3877-3878, 'R', 3879-3880, 'R', 3881-3882, 'R', 3883-3884, 'R', 3885-3886, 'R', 3887-3888, 'R', 3889-3890, 'R', 3891-3892, 'R', 3893-3894, 'R', 3895-3896, 'R', 3897-3898, 'R', 3899-3900, 'R', 3901-3902, 'R', 3903-3904, 'R', 3905-3906, 'R', 3907-3908, 'R', 3909-3910, 'R', 3911-3912, 'R', 3913-3914, 'R', 3915-3916, 'R', 3917-3918, 'R', 3919-3920, 'R', 3921-3922, 'R', 3923-3924, 'R', 3925-3926, 'R', 3927-3928, 'R', 3929-3930, 'R', 3931-3932, 'R', 3933-3934, 'R', 3935-3936, 'R', 3937-3938, 'R', 3939-3940, 'R', 3941-3942, 'R', 3943-3944, 'R', 3945-3946, 'R', 3947-3948, 'R', 3949-3950, 'R', 3951-3952, 'R', 3953-3954, 'R', 3955-3956, 'R', 3957-3958, 'R', 3959-3960, 'R', 3961-3962, 'R', 3963-3964, 'R', 3965-3966, 'R', 3967-3968, 'R', 3969-3970, 'R', 3971-3972, 'R', 3973-3974, 'R', 3975-3976, 'R', 3977-3978, 'R', 3979-3980, 'R', 3981-3982, 'R', 3983-3984, 'R', 3985-3986, 'R', 3987-3988, 'R', 3989-3990, 'R', 3991-3992, 'R', 3993-3994, 'R', 3995-3996, 'R', 3997-3998, 'R', 3999-4000, 'R', 4001-4002, 'R', 4003-4004, 'R', 4005-4006, 'R', 4007-4008, 'R', 4009-4010, 'R', 4011-4012, 'R', 4013-4014, 'R', 4015-4016, 'R', 4017-4018, 'R', 4019-4020, 'R', 4021-4022, 'R', 4023-4024, 'R', 4025-4026, 'R', 4027-4028, 'R', 4029-4030, 'R', 4031-4032, 'R', 4033-4034, 'R', 4035-4036, 'R', 4037-4038, 'R', 4039-4040, 'R', 4041-4042, 'R', 4043-4044, 'R', 4045-4046, 'R', 4047-4048, 'R', 4049-4050, 'R', 4051-4052, 'R', 4053-4054, 'R', 4055-4056, 'R', 4057-4058, 'R', 4059-4060, 'R', 4061-4062, 'R', 4063-4064, 'R', 4065-4066, 'R', 4067-4068, 'R', 4069-4070, 'R', 4071-4072, 'R', 4073-4074, 'R', 4075-4076, 'R', 4077-4078, 'R', 4079-4080, 'R', 4081-4082, 'R', 4083-4084, 'R', 4085-4086, 'R', 4087-4088, 'R', 4089-4090, 'R', 4091-4092, 'R', 4093-4094, 'R', 4095-4096, 'R', 4097-4098, 'R', 4099-4100, 'R', 4101-4102, 'R', 4103-4104, 'R', 4105-4106, 'R', 4107-4108, 'R', 4109-4110, 'R', 4111-4112, 'R', 4113-4114, 'R', 4115-4116, 'R', 4117-4118, 'R', 4119-4120, 'R', 4121-4122, 'R', 4123-4124, 'R', 4125-4126, 'R', 4127-4128, 'R', 4129-4130, 'R', 4131-4132, 'R', 4133-4134, 'R', 4135-4136, 'R', 4137-4138, 'R', 4139-4140, 'R', 4141-4142, 'R', 4143-4144, 'R', 4145-4146, 'R', 4147-4148, 'R', 4149-4150, 'R', 4151-4152, 'R', 4153-4154, 'R', 4155-4156, 'R', 4157-4158, 'R', 4159-4160, 'R', 4161-4162, 'R', 4163-4164, 'R', 4165-4166, 'R', 4167-4168, 'R', 4169-4170, 'R', 4171-4172, 'R', 4173-4174, 'R', 4175-4176, 'R', 4177-4178, 'R', 4179-4180, 'R', 4181-4182, 'R', 4183-4184, 'R', 4185-4186, 'R', 4187-4188, 'R', 4189-4190, 'R', 4191-4192, 'R', 4193-4194, 'R', 4195-4196, 'R', 4197-4198, 'R', 4199-4200, 'R', 4201-4202, 'R', 4203-4204, 'R', 4205-4206, 'R', 4207-4208, 'R', 4209-4210, 'R', 4211-4212, 'R', 4213-4214, 'R', 4215-4216, 'R', 4217-4218, 'R', 4219-4220, 'R', 4221-4222, 'R', 4223-4224, 'R', 4225-4226, 'R', 4227-4228, 'R', 4229-4230, 'R', 4231-4232, 'R', 4233-4234, 'R', 4235-4236, 'R', 4237-4238, 'R', 4239-4240, 'R', 4241-4242, 'R', 4243-4244, 'R', 4245-4246, 'R', 4247-4248, 'R', 4249-4250, 'R', 4251-4252, 'R', 4253-4254, 'R', 4255-4256, 'R', 4257-4258, 'R', 4259-4260, 'R', 4261-4262, 'R', 4263-4264, 'R', 4265-4266, 'R', 4267-4268, 'R', 4269-4270, 'R', 4271-4272, 'R', 4273-4274, 'R', 4275-4276, 'R', 4277-4278, 'R', 4279-4280, 'R', 4281-4282, 'R', 4283-4284, 'R', 4285-4286, 'R', 4287-4288, 'R', 4289-4290, 'R', 4291-4292, 'R', 4293-4294, 'R', 4295-4296, 'R', 4297-4298, 'R', 4299-4300, 'R', 4301-4302, 'R', 4303-4304, 'R', 4305-4306, 'R', 4307-4308, 'R', 4309-4310, 'R', 4311-4312, 'R', 4313-4314, 'R', 4315-4316, 'R', 4317-4318, 'R', 4319-4320, 'R', 4321-4322, 'R', 4323-4324, 'R', 4325-4326, 'R', 4327-4328, 'R', 4329-4330, 'R', 4331-4332, 'R', 4333-4334, 'R', 4335-4336, 'R', 4337-4338, 'R', 4339-4340, 'R', 4341-4342, 'R', 4343-4344, 'R', 4345-4346, 'R', 4347-4348, 'R', 4349-4350, 'R', 4351-4352, 'R', 4353-4354, 'R', 4355-4356, 'R', 4357-4358, 'R', 4359-4360, 'R', 4361-4362, 'R', 4363-4364, 'R', 4365-4366, 'R', 4367-4368, 'R', 4369-4370, 'R', 4371-4372, 'R', 4373-4374, 'R', 4375-4376, 'R', 4377-4378, 'R', 4379-4380, 'R', 4381-4382, 'R', 4383-4384, 'R', 4385-4386, 'R', 4387-4388, 'R', 4389-4390, 'R', 4391-4392, 'R', 4393-4394, 'R', 4395-4396, 'R', 4397-4398, 'R', 4399-4400, 'R', 4401-4402, 'R', 4403-4404, 'R', 4405-4406, 'R', 4407-4408, 'R', 4409-4410, 'R', 4411-4412, 'R', 4413-4414, 'R', 4415-4416, 'R', 4417-4418, 'R', 4419-4420, 'R', 4421-4422, 'R', 4423-4424, 'R', 4425-4426, 'R', 4427-4428, 'R', 4429-4430, 'R', 4431-4432, 'R', 4433-4434, 'R', 4435-4436, 'R', 4437-4438, 'R', 4439-4440, 'R', 4441-4442, 'R', 4443-4444, 'R', 4445-4446, 'R', 4447-4448, 'R', 4449-4450, 'R', 4451-4452, 'R', 4453-4454, 'R', 4455-4456, 'R', 4457-4458, 'R', 4459-4460, 'R', 4461-4462, 'R', 4463-4464, 'R', 4465-4466, 'R', 4467-4468, 'R', 4469-4470, 'R', 4471-4472, 'R', 4473-4474, 'R', 4475-4476, 'R', 4477-4478, 'R', 4479-4480, 'R', 4481-4482, 'R', 4483-4484, 'R', 4485-4486, 'R', 4487-4488, 'R', 4489-4490, 'R', 4491-4492, 'R', 4493-4494, 'R', 4495-4496, 'R', 4497-4498, 'R', 4499-4500, 'R', 4501-4502, 'R', 4503-4504, 'R', 4505-4506, 'R', 4507-4508, 'R', 4509-4510, 'R', 4511-4512, 'R', 4513-4514, 'R', 4515-4516, 'R', 4517-4518, 'R', 4519-4520, 'R', 4521-4522, 'R', 4523-4524, 'R', 4525-4526, 'R', 4527-4528, 'R', 4529-4530, 'R', 4531-4532, 'R', 4533-4534, 'R', 4535-4536, 'R', 4537-4538, 'R', 4539-4540, 'R', 4541-4542, 'R', 4543-4544, 'R', 4545-4546, 'R', 4547-4548, 'R', 4549-4550, 'R', 4551-4552, 'R', 4553-4554, 'R', 4555-4556, 'R', 4557-4558, 'R', 4559-4560, 'R', 4561-4562, 'R', 4563-4564, 'R', 4565-4566, 'R', 4567-4568, 'R', 4569-4570, 'R', 4571-4572, 'R', 4573-4574, 'R', 4575-4576, 'R', 4577-4578, 'R', 4579-4580, 'R', 4581-4582, 'R', 4583-4584, 'R', 4585-4586, 'R', 4587-4588, 'R', 4589-4590, 'R', 4591-4592, 'R', 4593-4594, 'R', 4595-4596, 'R', 4597-4598, 'R', 4599-4600, 'R', 4601-4602, 'R', 4603-4604, 'R', 4605-4606, 'R', 4607-4608, 'R', 4609-4610, 'R', 4611-4612, 'R', 4613-4614, 'R', 4615-4616, 'R', 4617-4618, 'R', 4619-4620, 'R', 4621-4622, 'R', 4623-4624, 'R', 4625-4626, 'R', 4627-4628, 'R', 4629-4630, 'R', 4631-4632, 'R', 4633-4634, 'R', 4635-4636, 'R', 4637-4638, 'R', 4639-4640, 'R', 4641-4642, 'R', 4643-4644, 'R', 4645-4646, 'R', 4647-4648, 'R', 4649-4650, 'R', 4651-4652, 'R', 4653-4654, 'R', 4655-4656, 'R', 4657-4658, 'R', 4659-4660, 'R', 4661-4662, 'R', 4663-4664, 'R', 4665-4666, 'R', 4667-4668, 'R', 4669-4670, 'R', 4671-4672, 'R', 4673-4674, 'R', 4675-4676, 'R', 4677-4678, 'R', 4679-4680, 'R', 4681-4682, 'R', 4683-4684, 'R', 4685-4686, 'R', 4687-4688, 'R', 4689-4690, 'R', 4691-4692, 'R', 4693-4694, 'R', 4695-4696, 'R', 4697-4698, 'R', 4699-4700, 'R', 4701-4702, 'R', 4703-4704, 'R', 4705-4706, 'R', 4707-4708, 'R', 4709-4710, 'R', 4711-4712, 'R', 4713-4714, 'R', 4715-4716, 'R', 4717-4718, 'R', 4719-4720, 'R', 4721-4722, 'R', 4723-4724, 'R', 4725-4726, 'R', 4727-4728, 'R', 4729-4730, 'R', 4731-4732, 'R', 4733-4734, 'R', 4735-4736, 'R', 4737-4738, 'R', 4739-4740, 'R', 4741-4742, 'R', 4743-4744, 'R', 4745-4746, 'R', 4747-4748, 'R', 4749-4750, 'R', 4751-4752, 'R', 4753-4754, 'R', 4755-4756, 'R', 4757-4758, 'R', 4759-4760, 'R', 4761-4762, 'R', 4763-4764, 'R', 4765-4766, 'R', 4767-4768, 'R', 4769-4770, 'R', 4771-4772, 'R', 4773-4774, 'R', 4775-4776, 'R', 4777-4778, 'R', 4779-4780, 'R', 4781-4782, 'R', 4783-4784, 'R', 4785-4786, 'R', 4787-4788, 'R', 4789-4790, 'R', 4791-4792, 'R', 4793-4794, 'R', 4795-4796, 'R', 4797-4798, 'R', 4799-4800, 'R', 4801-4802, 'R', 4803-4804, 'R', 4805-4806, 'R', 4807-4808, 'R', 4809-4810, 'R', 4811-4812, 'R', 4813-4814, 'R', 4815-4816, 'R', 4817-4818, 'R', 4819-4820, 'R', 4821-4822, 'R', 4823-4824, 'R', 4825-4826, 'R', 4827-4828, 'R', 4829-4830, 'R', 4831-4832, 'R', 4833-4834, 'R', 4835-4836, 'R', 4837-4838, 'R', 4839-4840, 'R', 4841-4842, 'R', 4843-4844, 'R', 4845-4846, 'R', 4847-4848, 'R', 4849-4850, 'R', 4851-4852, 'R', 4853-4854, 'R', 4855-4856, 'R', 4857-4858, 'R', 4859-4860, 'R', 4861-4862, 'R', 4863-4864, 'R', 4865-4866, 'R', 4867-4868, 'R', 4869-4870, 'R', 4871-4872, 'R', 4873-4874, 'R', 4875-4876, 'R', 4877-4878, 'R', 4879-4880, 'R', 4881-4882, 'R', 4883-4884, 'R', 4885-4886, 'R', 4887-4888, 'R', 4889-4890, 'R', 4891-4892, 'R', 4893-4894, 'R', 4895-4896, 'R', 4897-4898, 'R', 4899-4900, 'R', 4901-4902, 'R', 4903-4904, 'R', 4905-4906, 'R', 4907-4908, 'R', 4909-4910, 'R', 4911-4912, 'R', 4913-4914, 'R', 4915-4916, 'R', 4917-4918, 'R', 4919-4920, 'R', 4921-4922, 'R', 4923-4924, 'R', 4925-4926, 'R', 4927-4928, 'R', 4929-4930, 'R', 4931-4932, 'R', 4933-4934, 'R', 4935-4936, 'R', 4937-4938, 'R', 4939-4940, 'R', 4941-4942, 'R', 4943-4944, 'R', 4945-4946, 'R', 4947-4948, 'R', 4949-4950, 'R', 4951-4952, 'R', 4953-4954, 'R', 4955-4956, 'R', 4957-4958, 'R', 4959-4960, 'R', 4961-4962, 'R', 4963-4964, 'R', 4965-4966, 'R', 4967-4968, 'R', 4969-4970, 'R', 4971-4972, 'R', 4973-4974, 'R', 4975-4976, 'R', 4977-4978, 'R', 4979-4980, 'R', 4981-4982, 'R', 4983-4984, 'R', 4985-4986, 'R', 4987-4988, 'R', 4989-4990, 'R', 4991-4992, 'R', 4993-4994, 'R', 4995-4996, 'R', 4997-4998, 'R', 4999-5000, 'R', 5001-5002, 'R', 5003-5004, 'R', 5005-5006, 'R', 5007-5008, 'R', 5009-5010, 'R', 5011-5012, 'R', 5013-5014, 'R', 5015-5016, 'R', 5017-5018, 'R', 5019-5020, 'R', 5021-5022, 'R', 5023-5024, 'R', 5025-5026, 'R', 5027-5028, 'R', 5029-5030, 'R', 5031-5032, 'R', 5033-5034, 'R', 5035-5036, 'R', 5037-5038, 'R', 5039-5040, 'R', 5041-5042, 'R', 5043-5044, 'R', 5045-5046, 'R', 5047-5048, 'R', 5049-5050, 'R', 5051-5052, 'R', 5053-5054, 'R', 5055-5056, 'R', 5057-5058, 'R', 5059-5060, 'R', 5061-5062, 'R', 5063-5064, 'R', 5065-5066, 'R', 5067-5068, 'R', 5069-5070, 'R', 5071-5072, 'R', 5073-5074, 'R', 5075-5076, 'R', 5077-5078, 'R', 5079-5080, 'R', 5081-5082, 'R', 5083-5084, 'R', 5085-5086, 'R', 5087-5088, 'R', 5089-5090, 'R', 5091-5092, 'R', 5093-5094, 'R', 5095-5096, 'R', 5097-5098, 'R', 5099-5100, 'R', 5101-5102, 'R', 5103-5104, 'R', 5105-5106, 'R', 5107-5108, 'R', 5109-5110, 'R', 5111-5112, 'R', 5113-5114, 'R', 5115-5116, 'R', 5117-5118, 'R', 5119-5120, 'R', 5121-5122, 'R', 5123-5124, 'R', 5125-5126, 'R', 5127-5128, 'R', 5129-5130, 'R', 5131-5132, 'R', 5133-5134, 'R', 5135-5136, 'R', 5137-5138, 'R', 5139-5140, 'R', 5141-5142, 'R', 5143-5144, 'R', 5145-5146, 'R', 5147-5148, 'R', 5149-5150, 'R', 5151-5152, 'R', 5153-5154, 'R', 5155-5156, 'R', 5157-5158, 'R', 5159-5160, 'R', 5161-5162, 'R', 5163-5164, 'R', 5165-5166, 'R', 5167-5168, 'R', 5169-5170, 'R', 5171-5172, 'R', 5173-5174, 'R', 5175-5176, 'R', 5177-5178, 'R', 5179-5180, 'R', 5181-5182, '

Db	1908	-----PAKQIHGGILRLPAVEPTDQ-----	---AQY	1931
Qy	1108	IFRSHLEHQDLLKPSERRTSPVTLSPKHVSGFSSSLRTSTGDAGGSRPHRKPTIL		1167
		: :	: :	
Db	1932	LCRAH-----SSAGQOVARAVL-----HVUG-----GGPR-----		1957
Qy	1168	RKISAAQQLSASEVYVTHIGQFVALASGFLSVLLHCEALGHPRPTIISARNCEEEVOFSDRI		1227
		: :	: :	
Db	1958	-----VQVSPERTQVHAGRTVR-----LYCRAAGVPSATITWRKEGGSFPQARS		2002
Qy	1228	LLOPDDLSLIQALPEADVGYTCNATNALGYDYSIAVTL-----AGKPLVKTSRMTVIN		1282
		: :	: :	
Db	2003	ERTDIATILIPAITPADAGFYLCVATSAGVFAQRMQVVILSASDASPPGVK-----IE		2056
Qy	1283	TEKPAVTVDIGSTIKTVQGVNVTINCOVAGVPEAEVTFWRNKSGLGSPHHLHESLLTN		1342
		: :	: :	
Db	2057	SSSPSVT-----EGQILDLCNVVAGSAHAQVTVYRRGSLPPTHQVHGSRLRLPQ		2106
Qy	1343	VSSDQGLYSCHRAANLHGBELTSTQILLIDPPQVPTQLEDIRALLAAVCPNLPVLTSP		1402
		: :	: :	
Db	2107	VSPADSGEYVCRVENGSGPKASITVSLHGTHT-----SGPSITPV-----PG		2149
Qy	1403	GTQLV-LDPGNSAL-----LGCPKIGHVPEVNIWTF-HGGQPIVTTATGLTHHILAAQGI		1453
		: :	: :	
Db	2150	STPIRIEYSSSHRVARGQTLDLCNVVPGQAAQVTHWKRKGS--LPARHQTH-----GSL		2202
Qy	1454	LOVANLSGGSQGEFFSCLAQNEAGVLMOKASLVI		1486
		: :	: :	
Db	2203	LRHQVTPADSGEYVCHVGTSGPL--EASVLV		2233
RESULT 14				
T20992				
hypothetical protein F15G9.4a - Caenorhabditis elegans				
C:Species: Caenorhabditis elegans				
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000				
C:Accession: T20992; T24733				
R:Submitted to the EMBL Data Library, December 1994				
A:Reference number: Z19355				
A:Accession: T20992				
A:Status: preliminary; translated from GB/EMBL/DBJ				
A:Molecule type: DNA				
A:Residues: 1-5175 <WIL>				
A:Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a				
A:Experimental source: clone F15G9				
R:Kershaw, J.				
Submitted to the EMBL Data Library, December 1994				
A:Reference number: Z19929				
A:Accession: T24733				
A:Status: preliminary; translated from GB/EMBL/DBJ				
A:Molecule type: DNA				
A:Residues: 1-5175 <W12>				
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a				
A:Experimental source: clone T09B9				
C:Genetics:				
A:Gene: CESP:F15G9.4a				
A:Map position: X				
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/1; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1				
Query Match 3.7%; Score 352.5; DB 2; Length 5175;				
Best Local Similarity 19.1%; Pred. No. 2.3e-11;				
Matches 188; Conservative 138; Mismatches 329; Indels 331; Gaps 388				
Qy	897	LKCPARRVRKPLITWEKDGCHLISSTHVIVAPFGYVKLHRLKPSDAGVYTC-----SAGPA		952
		: :	: :	
Db	3146	IRCEVFGNPEVNNLKGDEPTSDLLQFSTKLSYLHRETTIADGGYTCIATNKAGES		3205
Qy	953	REHVFILKLGGRKIVARPLSPR-----SEEEVLAKRG-----GPREALQTHKHONG		1000
		: :	: :	

Db 3206 QTTTDEVEVL-----VPRTEDEERVLOQKGNVYMHVCQVTPVYVTKR----- 3252  
QY 1001 IFSNGSKAEK-----RGLAANPGSRYDDLVSLRLLGOGWPGPELLASWEAQDSARNIT 1053  
Db 3253 ---NGKEIQPNVPLHNRATRADEKGYSCIASN--EAGTAVADFLLDVFTKPTFETHET 3307  
QY 1054 SEEDPGAEQVLLHL-----PFT-----MVTEORRLDDILG 1083  
Db 3308 TFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMNIIILSPRGDTLMILKAQRFD---G 3364  
QY 1084 NL-----SQPEELRDLYSKHLVAQLAQEIFRSHLEHODTLTKPSERTSPVTL 1132  
Db 3365 GLYTCVATNSYGDSEQDFKV--NVYTKPIDETIDQTPKAVAGEIILKCPVLGNPTPTVT 3423  
QY 1133 -----SPKHVGSFSSLSRTSSTDAG-----GSRRPD-----RKPTI 1166  
Db 3424 WKRGDDAVPNDSRHTVNNYDLKINSVTEDAGYSCIADVNEAGNLTHYAAEVIGKPTF 3483  
QY 1167 LRKISAAQQLSASEVTHLQGTVALASGTLVLLHCEAIGHPRPTTISWARGVEVOFSDR 1226  
Db 3484 VRKGGNLYEVIENDIT-----MDCGVTSRPLPSISWFRGDKPVLVYDR 3527  
QY 1227 ILLOPDDS--LOILAPVEADVGYTCNATNALGYDSVSIATVLAGKPLVKTSRMTVINTEK 1285  
Db 3528 YSISPDGSHITINKAKLSGDKGYICRASNEAGTSDIDLILKLVPPKIDKSN----- 3580  
QY 1286 PAVTVDIGSTIKTVQGVNVTINCOVAGVPEAEVTFWRNKSGLSPHHLHGSLLT----- 1341  
Db 3581 -----IGNPLAIV--ARTIYLECPISGIPQDPVITWTKN-----GMDINMTDSRVILAQNE 3629  
QY 1342 -----NVSSDDQGLYSYCRANLHGLTETQILLIDPPO-----V 1376  
Db 3630 TFGIENVVOTDQGRVTCATNRGGRASHDFSLDVLSPPEFDIHGQPTIKREGDTITLTC 3689  
QY 1377 PQQL-----EDIRAL-----LAATGPNL-----PSVLTSPLGTOLVLD 1409  
Db 3690 PIKLAEDIAOVMOVSWTKDSRALDGLTDNDVDSDDGRKLTISQASLENAGLYTCIALN 3749  
QY 1410 PGNSA-----LGCPIKHPVPNITWFHGGQPTV 1438  
Db 3750 RAGEASLEFKVEILSPVIDISRNDVQPVAVNQPTIMRCVATGTHFFPSIKLWKGKEVT 3809  
QY 1439 TATGLTHHILAAQTLQVANI--SGSQGFSCIAQNEAGVLMOKASILVLDQYVWMSVDRL-- 1496  
Db 3810 DDENI--RIVEQGOVLQILRTSDSHAGKWSVVAENDAGV--KELEWLDVFTPPVVSWS 3865  
QY 1497 -----ATCSAGNGVQVQPLRLCLLNSTEVPNAPCAKVRPAVQPIACNRDRCP-- 1546  
Db 3866 DNPICALGETITLFCNAGSGNYPOLKWKAGGSLIFDPSDGARI-----SLKAGRLDIPHL 3920  
QY 1547 SRMVTWSACTRSCGGGVQTRRTVCOKLKASGISTPVSNDMCTQVAKRPVDTQACNOQL 1606  
Db 3921 KKTVDGY-----TCQALNAG-----TSEASVSVDV----- 3947  
QY 1607 CVWEAFSSWGQCNPCGPHLAVOHQVFCOTRDI-----TLPSE-----OCSALPRPVS 1657  
Db 3948 -----LVPEET-----NRDGDMSPLPQAQSLTLOCLAQGKPPV 3982  
QY 1658 TQNCWSEACSVHVRVSLWTLCTATCG 1683  
Db 3983 -----QMRWTLNGTALHTHSTPG 3999

RESULT 15  
T43290  
hemiscentin precursor - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T43290; T20993; T24734  
R:Vogel, B.E.; Hedgecock, E.M.  
submitted to the EMBL Data Library, June 1998  
A:Description: Hemiscentin is required for hemidesmosome mediated cell adhesion and germ-  
A:Reference number: 222396

A:Accession: T43290  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5198 <VOG>  
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1  
R:Sulston, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19355  
A:Accession: T20993  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <WIL>  
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b  
R:Xershaw, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24734  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <W12>  
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b  
A:Experimental source: clone T09B9  
C:Genetics:  
A:Gene: him-4; F15G9.4b  
A:Map position: X  
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184  
; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 303  
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

Query Match 3.7%; Score 352.5; DB 2; Length 5198;  
Best Local Similarity 19.1%; Pred. No. 2.3e-11;  
Matches 188; Conservative 138; Mismatches 329; Indels 331; Gaps 38;  
QY 897 LRCPARRRKPKITWCKDGOHLISSTHTVAPPGYLKIHLKPSDAGVTC-----SAGPA 952  
Db 3146 IRCEVGEFNPVNWLLKDGEPYISDQLQSTKLSYLHLRETTLADGGTTCATNKAGES 3205  
QY 953 REHFVTKLGGNKLVARPLSPR--SEEVLAGRKG-----GPKALQTHQNG 1000  
Db 3206 QTTTDEVEVL-----VPRTEDEERVLOQKGNVYMHVCQVTPVYVTKR----- 3252  
QY 1001 IFSNGSKAEK-----RGLAANPGSRYDDLVSLRLLGOGWPGPELLASWEAQDSARNIT 1053  
Db 3253 ---NGKEIQPNVPLHNRATRADEKGYSCIASN--EAGTAVADFLLDVFTKPTFETHET 3307  
QY 1054 SEEDPGAEQVLLHL-----PFT-----MVTEORRLDDILG 1083  
Db 3308 TFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMNIIILSPRGDTLMILKAQRFD---G 3364  
QY 1084 NL-----SQPEELRDLYSKHLVAQLAQEIFRSHLEHODTLTKPSERTSPVTL 1132  
Db 3365 GLYTCVATNSYGDSEQDFKV--NVYTKPIDETIDQTPKAVAGEIILKCPVLGNPTPTVT 3423  
QY 1133 -----SPKHVGSFSSLSRTSSTDAG-----GSRRPD-----RKPTI 1166  
Db 3424 WKRGDDAVPNDSRHTVNNYDLKINSVTEDAGYSCIADVNEAGNLTHYAAEVIGKPTF 3483  
QY 1167 LRKISAAQQLSASEVTHLQGTVALASGTLVLLHCEAIGHPRPTTISWARGVEVOFSDR 1226  
Db 3484 VRKGGNLYEVIENDIT-----MDCGVTSRPLPSISWFRGDKPVLVYDR 3527  
QY 1227 ILLOPDDS--LOILAPVEADVGYTCNATNALGYDSVSIATVLAGKPLVKTSRMTVINTEK 1285  
Db 3528 YSISPDGSHITINKAKLSGDKGYICRASNEAGTSDIDLILKLVPPKIDKSN----- 3580  
QY 1286 PAVTVDIGSTIKTVQGVNVTINCOVAGVPEAEVTFWRNKSGLSPHHLHGSLLT----- 1341  
Db 3581 -----IGNPLAIV--ARTIYLECPISGIPQDPVITWTKN-----GMDINMTDSRVILAQNE 3629  
QY 1342 -----NVSSDDQGLYSYCRANLHGLTETQILLIDPPO-----V 1376  
Db 1376



Query Match	1.1%	Score 56.4;	DB 1;	Length 2277;
Best Local Similarity	44.9%;	Pred No. 0.00041;		
Matches 256;	Conservative	0;	Mismatches 311;	Indels 3; Gaps 1;
2936 aggtgcttgcggggaggaaggcgccgcgaagagggccctgcagcccaaacaccaga				2995
530 AGGAGATCAAGGGCGCACCGTGCCGTGGACTGGCCGCTGACAGG				589
2996 acgggatcttccaaacggcgagcaaggcgggaagcggggcgctggccgcaaccgggga				3055

```

; MOLECULE TYPE: CDNA
US-08-676-974-5

Query Match      1.1%; Score 56.4; DB 1: Length 2277;
Best Local Similarity 44.9%; Pred. NO. 0.00041;
Matches 256; Conservative 0; Mismatches 311; Indels 3; Gaps

QY 2936 aggtgcttgccgggaggaagggcgccgcgaagagagccctgcagaccacaacaccaga 2995
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 530 AGGAGATCAAGGGCCGCGCCGCTGGCGCTGGACTGGCCCTGCCCRAGGACAGTACAAG 589
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2996 acgggatcttccaaacgcagcaagcgaggaagcgggcgctggccgcgaaccccgggga 3055
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 590 ACACCCAGAGCGTGAGCGCCATCGCGGAGGAGAAGAGCCACGAGAGCAAGCACGAGAGA 649
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3056 gccgctacgaacactctgtctcccgctgctgagcagaggcgctgcccggagagctgc 3115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 650 GCCTGAAGAAGAGGCCGCCGAGGAGGAGACATGCGAGGAGGAGGACGACGACGACGACG 709
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3116 tgacctgtggaggcgaggactccgcggaaggaaacacgacctcgagaggaggaaccgg 3175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 710 ACGAGCAGACGACGAGGAGGACGCGCTGTTCCACGACGAGGACGAGGAGGAGGAACA 769
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3176 gtgcagagcaagtgtctctgcacttgccttcaccatggttaccgcgagcgcgccctgg 3235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 770 TCGAGAGCAAGGTGACCAAGCCGCTGCAGATCCAGAAGCGCGCGTGAAGCGCCCGCC 829
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3236 acgacatctctgggaaacctctccagcagcccgaggagctgcgcacctctacagcaagc 3295
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 830 CGGCCAAGACGACGCGACCAACGCGAGGAGGACAGCACCTTGGAGGAGCGACGACATCG 889
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3296 acctgtgtgccagcttgcccagagagatcttcgcgagccacttgagacaccaggaacagc 3355
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 890 ACAGACGGGAGGAGCTGGCC---AGAGCGACACCACGACCGAGGACGAGGAGCAAGG 946
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3356 tcttgaagcctcggagcgaggacttcccagtgactctctcgtctataaacacgtgt 3415
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 947 CCGTGCAGGTGAGCAACAAGAAGCGCAAGCTGCCACGCGCTGAACGAGGGCGAAGA 1006
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3416 ctggttcaacgaactcctctcggaacctctccaccgggagcgccggggaggtctctgaa 3475
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1007 CCGTGTTCATCCGCAACTGAGCTTCACACCGAGGAGGAGGAGGACCTGGCGAGCTGCTGC 1066
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3476 ggccaacccgcaagcccaccatctctgcga 3505
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1067 AGCAGTTCGCGAGCTGAAGTACGTGGCGCA 1096
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

3
RESULT"
US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Osman Ph.D., Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UCB96-055
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415)343-4341
: TELEFAX: (415)343-4342
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2277 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-09-098-487-5

```

Query Match 1.1%; Score 56.4; DB 2: Length 2277;  
Best Local Similarity 44.9%; Pred. No. 0.00041;  
Matches 256; Conservative 0; Mismatches 311; Indels

Matches	256; Conservative	0; Mismatches	311; Indels	3; Gaps
QY	2936	aggtgcttcgycggaggaagggcggcccgaaagagagcccttcagaccacacaaaccaga	2935	
Db	530	AGGAGATCAGGGCCGACCGTGGCCGTGGACTGGGCCGTGGCCCAAGGACAAGTACAAGG	589	
QY	2996	acggggatcttctcaacggcagcaaggcgagagcgggccctggccgcaaccgccgggga	3055	
Db	590	ACACCAGAGCGTGAGCGCCATCGGCGAGGAGAGACCCACGAGAGCAAGCACCCAGAGA	649	
QY	3056	gcgcgtacagacaccttcgtctccggctgtctgtagcaggcggtgtgcccggagagctgc	3115	
Db	650	GGTGAGAGAGAGGGCCGCGAGGAGGAGGCACATGGAGGAGGAGGAGACGACGACGACG	709	
QY	3116	tggccctcgtgggaaggcgaggaactccgcgggaagaagcaacgcacctcgagagagacccgg	3175	
Db	710	AGCAGACGACGACGAGGAGGAGCGCGTGTTCGACGACGAGGACGAGGAGGAGGAGACA	769	
QY	3176	gtgcagagcaagtgtccttcgcacctgcccttcacctgtgtgaccgagcagcgcgcctgg	3235	
Db	770	TCGAGAGCAAGGTGACCAAGCCGTCGATCCAGAAGCGCGCCGCTGAAGCGCCCGCCC	829	
QY	3236	acgacatcctggggaaacctctccagcagcccgagagctgcgcgacctctacagcaagc	3295	
Db	830	CGGCCAAGAGCAGCGACCAACAGCGAGGAGGACAGCGACCTGGAGGAGAGCGACAGCATCG	889	
QY	3296	acctggtggcccgactgcccagagagattctccgagccacctggagcaccagcacgc	3355	
Db	890	ACGACGGCGAGAGTGTGCC---AGAGCGACACAGCACCGGAGGAGCAGGAGGACAAGG	946	
QY	3356	tctgaagccctcggagcgcaaggacttcccagtaactcttcgctctataaacagtgt	3415	
Db	947	CGGTCAAGTGAGCAACAAAGAGAAGCGCAAGCTGCCACGCGACGTGAACGAGGGCAAGA	1006	
Y	3416	ctggcttcagcagctccctcggaacctctccaccgaggagcccgggaggctctcgaa	3475	
b	1007	CCGTGTTTCATCCGAACCTGAGCTTCGACAGCGAGGAGGAGGAGCTGGCCGAGCTGCTGC	1066	
Y	3476	ggccacaccggaagcccaacctctctgcga	3505	
b	1067	AGCAGTTGGCGAGCTGAAGTACGTGCCCA	1096	

## RESULT AND

```

RESOLUTION 4
US-08-125-468-1
; Sequence 1, Application US/08125468
; Patent No. 5589385
;
; GENERAL INFORMATION:
;
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantini, Susan E.
;
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmids

```

```

; TITLE OF INVENTION: useful therein
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
;
; ZIP: 07470
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/125,468
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; PS-08-125-468-1

```

Query Match 1.0%; Score 54.6; DB 1; Length 30001;  
Best Local Similarity 45.6%; Pred. No. 0.0029;  
Matches 235; Conservative 0; Mismatches 274. Indexes 6.

QY	2960	gocgaaaggagccctgcagaccacaacacccagaaaggatcttctccacgcagca	3019
Db	15355	GC CGATGCTGC CCGCTCAACACCGCTACTTCCAGCGCAGAGCGGTGCG	15414
QY	3020	aggcggaagaagcgggctgcgcaccaaccgggagcgctaaagacacctcgtctccc	3079
Db	15415	AGCTCGGCGCTTTCAGCTGCTGCGCGAGCGCTCCGCCGGCTGCACCATCCGCGCG	15474
QY	3080	ggctgctggagcagggcggtggccggagagctgctggtcctcgtgggagcgaggaact	3139
Db	15475	AAC TGGGCGTCCG CACCGCTGITCANGACTTCCTGAAGCCCTGTCGCGCTCGGCC	15534
QY	3140	ccgcggaagaacacagacctcgaggagaccgggtgcagagacaagtgcctcgcacc	3199
Db	15535	TGCTGGACGAGCAGGACGGCGGCTACCGGGGCTCCGAGCTGCCCGGAGTTCCTGCTCC	15594
QY	3200	tgccttcaccatgtagccagagcagcgcgctcgacgacatcctgggaacctctcc-	3258
Db	15595	CGGCCCCACGTACCTCGGGGGCACCCCGCCACGACGCTCGGCTGCATACCAACGCCT	15654
QY	3259	-----cagcagcccgaggagctgcgcagacctctacagcaagcacctggtggccagctgg	3313
Db	15655	GGCGCGAGCTGACCGACGCGCTGCGCAGCGCAAGGCCAAGTCGGCCGTGGCCGCGCAGG	15714
QY	3314	ccagagagatcttcgcagccacctggagcaccagacacgctctctgaagccctcgagc	3373
Db	15715	GCCAGCTGGCCTACCCCAAGACATACGAGGACCTGGACCGCGCCGCGCAGATCATGCTGC	15774
QY	3374	gcaggacttccccagtgactctctgcctcataaacacgtgtctggttcagcagctccc	3433
Db	15775	ACATGACGCCCAACAGGTTTCACGGCCGACGATGCGCGCGCGATCGACTGGAGCC	15834
QY	3434	tgcggacctctccaccgaggacgcggggagcg	3468





## RESULTS

```

; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickinson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; COMPANY: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,004
; FILING DATE: 27-JUL-1995
; CLASSIFICATION: 435
; PRIORITY DATA:
; PRIOR APPLICATION NUMBER: US 08/402,282
; FILING DATE: 10-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"

```



	LOCATION:	6403..7770	"potential open reading"
	OTHER INFORMATION:	/function=	
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	7770..8006	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	8033..8236	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	8244..9443	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	9450..10244	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	10371..10586	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	11115..11786	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	11917..12741	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	12748..14499	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	14771..15154	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	15154..15426	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		
	NAME/KEY:	misc_feature	
	LOCATION:	15429..15664	
	OTHER INFORMATION:	/function=	"potential open reading"
	OTHER INFORMATION:	frame"	
	FEATURE:		

	Query Match	1.0%	Score 53;	DB 1;	Length 15664;
	Best Local Similarity	49.5%	Pred. No. 0.0055;		
	Matches 137;	Conservative 0;	Mismatches 140;	Indels 0;	Gaps 0;
Qy 3075	ctcccggtctgtgagcaaggcgtggcccgagagctgctgacctggtggagggcgca	3134			
Db 12078	cccggtggacactgtctgttcgacgcctgtgcccgcacactgttctgcattgtggcggcgca	12137			
Qy 3135	ggactccggaaagaaacacgacctcgagaggagaccgcgggttcagagcagaagtgtcctc	3194			
Db 12138	cgagtttcacaaacggcagggcgccgacccggcagggagacaccttggacggcggtgatttcca	12197			





```

US-08-863-010-4
; Sequence 4, Application US/08863010
; Patent No. 6087146
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,010
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,002
; FILING DATE: 29-SEP-1995
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP No. 6087146 yet received
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2889 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PS-08-863-010-4

```

[illegible]

```

Db 1726 G-----TGGACACCCCTGGTCCACGAAAGGGGCGGAGGAGCTCCTAAACGCGCCTC 1777
QY 3256 tccagcagcccgagagagctgcgcagactctacagcaagcactctc---ggtggcccgagctg 3312
Db 1777 GCCACACCCCTGAAGGAGAGAGCTGGCTCGCCCTCAAGCCCGCAGAAGGTGGCCCTCCCTG 1836
QY 3313 gccagggagatctccgcagccagccactctggagcaccagggacacgctctctgaagccctcgag 3372
Db 1837 GAGCGCCCTCCGCTTCAGAAAGACCCGCCCTTTACCTCACCTGCTCCAGCTGGAGAAC 1896
QY 3373 gcacgaacttcccgagtgactctctcgc 3400
Db 1897 CACAGGACCCCTCCAGGCTCTCCCTCCCC 1924

RESULT 13
US-09-024-429-4
; Sequence 4, Application US/09024429
; Patent No. 6165768
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,429
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,002
; FILING DATE: 29-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 255829/1995
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: TSUSAKI=1b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2889 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-024-429-4

```

Query Match 1.0%; Score 50.4; DB 4; Length 2889;  
Best Local Similarity 48.0%; Pred. No. 0.012;  
Matches 215; Conservative 0; Mismatches 221; Indels 12









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 22:34:46 ; Search time 3724.92 Seconds  
(without alignments)  
19164.276 Million cell updates/sec

Title: US-10-044-807-1  
Perfect score: 5289  
Sequence: 1 atggaatgctgcgtggc.....gaacttggcgaagcgtga 5289

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	824	15.6	1009	10	BM476141	BM476141 AGENCOURT
2	623.2	11.8	850	10	BE888902	BE888902 601513828
3	562.2	10.6	568	9	A1917724	A1917724 t11c08 x
4	521	9.9	522	9	A1342006	A1342006 qs98d10 x
5	514.4	9.7	581	10	BT523618	BT523618 603051757
6	508	9.6	509	9	A1261611	A1261611 qz35h01 x
7	502	9.5	622	9	B8633755	B8633755 BB633755
8	478	9.0	979	11	AK020115	AK020115 Mus muscu
9	471.8	8.9	633	9	BB620571	BB620571 BB620571
10	456.8	8.6	660	9	BB659292	BB659292 BB659292
11	443	8.4	700	9	BB643318	BB643318 BB643318
12	442.6	8.4	449	10	BF058634	BF058634 7k24d09 x
13	414.8	7.8	419	9	A1750970	A1750970 cn06h02 y
14	413.8	7.8	619	9	BB478282	BB478282 BB478282
15	409.4	7.7	555	9	BB615232	BB615232 BB615232
16	407.6	7.7	637	9	BB612331	BB612331 BB612331
17	406.2	7.7	586	10	BF510853	BF510853 UI-H-B14-

c	18	398.8	7.5	454	10	N33407	N33407 yy41f11.s1
19	386.2	7.3	550	10	BE628435	BE628435 uu09f09.y	
20	384.4	7.3	643	9	BB193444	BB193444 BB193444	
c	21	380	7.2	380	9	AI825646	AI825646 wb75b10.x
c	22	375.4	7.1	644	10	BF111214	BF111214 7n4e07.x
c	23	365.8	6.9	423	9	AA482392	AA482392 zv05h07.r
c	24	350.4	6.6	631	10	BF445136	BF445136 nad2la04.
c	25	338	6.4	600	10	BE631939	BE631939 uu09f09.x
c	26	327.8	6.2	628	9	AI459225	AI459225 tk11b03.x
27	288	5.4	846	11	AK020361	AK020361 Mus muscu	
28	285	5.4	313	10	W47220	W47220 zc39a07.r1	
c	29	277.6	5.2	539	9	AW120520	AW120520 UI-M-BH1-
30	271	5.1	442	9	B8839998	B8839998 BB839998	
31	268	5.1	606	10	BG750735	BG750735 602706924	
c	32	266	5.0	452	9	AW120794	AW120794 UI-M-BH2.
c	33	264.2	5.0	485	9	AA707140	AA707140 zj33e01.s
c	34	263.2	5.0	302	9	AA284278	AA284278 zc39a07.T
c	35	247.2	4.7	437	10	W47029	W47029 zc39a07.s1
c	36	238.2	4.5	534	9	AA971699	AA971699 op95a12.s
37	232	4.4	233	10	N44869	N44869 yy40f11.r1	
c	38	217	4.1	223	10	BE501165	BE501165 7a39b04.x
c	39	206	3.9	742	12	AG029671	AG029671 Pan trogl
40	200.4	3.8	425	10	BF767952	BF767952 CMI-CN006	
41	187	3.5	444	12	AQ354354	AQ354354 CITBI-El-	
42	174.8	3.3	605	9	AI905640	AI905640 CM-BT094-	
c	43	174.4	3.3	332	9	AV161151	AV161151 AV161151
44	164	3.1	164	10	BF932251	BF932251 IL2-NW019	
45	160	3.0	571	9	AI905601	AI905601 CM-BT094-	

#### ALIGNMENTS

BM476141 1009 bp mRNA linear EST 05-FEB-2002  
AGENCOURT\_647885 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:558669  
5', mRNA sequence.

BM476141  
EST.

BM476141.1 GI:18525183  
human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1009)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LLAM12282 row: a column: 06

High quality sequence stop: 647.

Location/Qualifiers

1..1009

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:558669"

/clone.lib="NIH\_MGC\_88"

/tissue.type="duodenal adenocarcinoma, cell line"

/lab\_host="BHI08 (phage-resistant)"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site 1: NotI; Site 2: SalI; Cloned unidirectionally;

oligo-dt primed. Average insert size 1.767 kb. Library

enriched for full-length clones and constructed by Life

Technologies. Note: this is a NIH\_MGC Library."

FEATURES  
source



QY 361 ctcaagtgcccaagcacaagcaacccctggtgttggaactagacacctaaggtcttagat 420  
 Db 404 CTCAGTCCCAAGCCAAAGGAACAACCCCTGGTGTGTGAACATAGCACCTTAAGGTCTTAGAT 463  
 QY 421 ggtacgogtctatcacagaatctttgatatgtgcacacagtggttttatgccaattgtt 480  
 Db 464 GGTACGGC-TGCTATACAGATCTTTGGATATGTGCATCAGTGGTTTATGCCAAATTTGT 522  
 QY 481 ggtcgatcacagcagcgggaagcaccgtcaaggaataactgtgggtctgcaacgga 540  
 Db 523 GGCTGCGATCACCAAGCTGGGAAGCAGCCGTCAAGGAAGATAACTGTGGGTCTGCAACGGA 582  
 QY 541 gat-ggggtccacctgcggtgtccagagggcagtgatataatcccagctctccgcaacaa 599  
 Db 583 GATGGGTCCCACTGCGCGTGGTCCGAGGCGAGTATAATCCCAAGCTCTCAGCAACCCA 642  
 QY 600 atc-ggatgatactgtgtgtgc-aattccctatggaagttagacatatcgcttgttelta 657  
 Db 643 ATCGGGATGATACTGTGTGTGCAATTCCTATGGGAAGTTGACTATTTCGCTTGGTTTA 702  
 QY 658 aaaggtcctgatcacttatctctggaacacaaacccctccaggggactaaaggtgaaac 717  
 Db 703 AAAGGGCTGATCATTCTTTATTGGGAACCAAAA-CTCCACGGGACTCAAGGTGGCAAC 761  
 QY 718 agtctcagctccacag 733  
 Db 762 AGGTTCTATCCAGAG 777

RESULT 3  
 A1917724/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 HOMO sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 568)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/Image.html](http://www-bio.llnl.gov/bbrp/image/Image.html)  
 Insert length: 922 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 437.  
 Location/Qualifiers  
 1..568  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2240462"  
 /clone\_lib="NCI-CGAP GC6"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; plasmid DNA  
 from the normalized library NCI-CGAP GC4 was prepared, and  
 ss circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clones IDs  
 1257096-1258631, 1469064-1470983, and 1475592-1476743).  
 Subtraction by Bento Soares and M. Fatima Bonaldo.  
 BASE COUNT 120 a 157 c 169 g 121 t 1 others  
 ORIGIN

Query Match 10.6%; Score 562.2; DB 9; Length 568;  
 Best Local Similarity 99.3%; Pred. No. 3.3e-105;  
 Matches 564; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1996 tgccacgcaagtgaggaaattggcaagtggagtcacatgtagtcacatgagggtcgcg 2055  
 Db 568 TGCCACGCAAGGTGGGANATTTGGCAAGTGGAGTCCATGTCTCACATGTGCGGTGGCG 509  
 QY 2056 ctacagaccagagacgtctctcgcagccacctgtttccagagagatgaatgaacagtc 2115  
 Db 508 CTACAGACCAGAGACGTCTTCTGCAGCCACCTGCTGCCAGAGAGATGAATGAACAGTC 449  
 QY 2116 atcctgctgatgagctgtgtcgcagcccaagcccaagccagcagctgtgcaagctgttaacgc 2175  
 Db 448 ATCCTGCTGATGAGCTGTGTGCCAGCCCAAGCCAGCAGCGGTGCAAGCTTGTAAACCGC 389  
 QY 2176 tttaattgccccccagcctggtacccctgacagtgccagcgtgttccagaacgtgtggc 2235  
 Db 388 TTTAATTGCCCCCCAGCCTGTGTACCCCTGCACAGTGGCAGCGCTGTTCACAGAACGTGTGGC 329  
 QY 2236 gggggtgttcagaaacgtgaggtcttcttgcagcagcgcagctggtgatgagcagcttctg 2295  
 Db 328 GGGGGTGTTCAGAAACGTGAGGTCTTTTGAAGCAGCGCATGGCTGATGGCAGCTTCTCTG 269  
 QY 2296 gagcttctgagaccttctgttcagcttcaaaacctgctgcccagcagcagcagcagcagc 2355  
 Db 268 GAGCTTCTGAGACCTTCTGTTCAGCTTCAAACTTCCCTGCCAGCAGCAGCATGCAAGAA 209  
 QY 2356 gatgactgtccagcagcagtggtctctctcagactgacagctgttccacaagctgcggg 2415  
 Db 208 GATGACTGTCCACGAGTGGCTTCTCTCAGACTGGACAGAGTGTCCACAGCTGCGGG 149  
 QY 2416 gaaggcaccacagactcgaagcgcattttccgaaagatgctgaaacccgctctcaagc 2475  
 Db 148 GAAGGCACCAGACTCGAAGCGCCATTTCGCGAAAGATGCTGAAACCCGCGCTCTCAACG 89  
 QY 2476 gttgtcaattccacctgtgcccgccttctctctctctctctctctctctctctctctct 2535  
 Db 88 GTGTCAATTCCACCTGTGCGCGCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 29  
 QY 2536 ctgggaacctgtgcaaggcccgccggcgcc 2563  
 Db 28 CTGGCAACCTGTGCAAGGCACGCGCGC 1

RESULT 4  
 A1342006/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 522)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

BI523618 581 bp mRNA linear EST 29-AUG-2001  
603051757F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5201199 5',  
mRNA sequence.

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Db 366



RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome*. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

	FEATURES	SOURCE
1.	100% cotton	100%
2.	100% polyester	100%
3.	100% rayon	100%
4.	100% silk	100%
5.	100% wool	100%
6.	100% linen	100%
7.	100% leather	100%
8.	100% rubber	100%
9.	100% metal	100%
10.	100% plastic	100%
11.	100% glass	100%
12.	100% paper	100%
13.	100% wood	100%
14.	100% stone	100%
15.	100% concrete	100%
16.	100% brick	100%
17.	100% tile	100%
18.	100% asphalt	100%
19.	100% gravel	100%
20.	100% sand	100%
21.	100% dirt	100%
22.	100% mud	100%
23.	100% water	100%
24.	100% air	100%
25.	100% fire	100%
26.	100% lightning	100%
27.	100% thunder	100%
28.	100% rain	100%
29.	100% snow	100%
30.	100% ice	100%
31.	100% frost	100%
32.	100% hail	100%
33.	100% sleet	100%
34.	100% drizzle	100%
35.	100% fog	100%
36.	100% mist	100%
37.	100% dew	100%
38.	100% frost	100%
39.	100% snow	100%
40.	100% ice	100%
41.	100% frost	100%
42.	100% snow	100%
43.	100% ice	100%
44.	100% frost	100%
45.	100% snow	100%
46.	100% ice	100%
47.	100% frost	100%
48.	100% snow	100%
49.	100% ice	100%
50.	100% frost	100%
51.	100% snow	100%
52.	100% ice	100%
53.	100% frost	100%
54.	100% snow	100%
55.	100% ice	100%
56.	100% frost	100%
57.	100% snow	100%
58.	100% ice	100%
59.	100% frost	100%
60.	100% snow	100%
61.	100% ice	100%
62.	100% frost	100%
63.	100% snow	100%
64.	100% ice	100%
65.	100% frost	100%
66.	100% snow	100%
67.	100% ice	100%
68.	100% frost	100%
69.	100% snow	100%
70.	100% ice	100%
71.	100% frost	100%
72.	100% snow	100%
73.	100% ice	100%
74.	100% frost	100%
75.	100% snow	100%
76.	100% ice	100%
77.	100% frost	100%
78.	100% snow	100%
79.	100% ice	100%
80.	100% frost	100%
81.	100% snow	100%
82.	100% ice	100%
83.	100% frost	100%
84.	100% snow	100%
85.	100% ice	100%
86.	100% frost	100%
87.	100% snow	100%
88.	100% ice	100%
89.	100% frost	100%
90.	100% snow	100%
91.	100% ice	100%
92.	100% frost	100%
93.	100% snow	100%
94.	100% ice	100%
95.	100% frost	100%
96.	100% snow	100%
97.	100% ice	100%
98.	100% frost	100%
99.	100% snow	100%
100.	100% ice	100%

```

Location/Qualifiers
1. .622
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="A330077F03"
  /clone_lib="RIKEN full-length enriched, adult male spinal
  cord"
  /sex="male"
  /tissue_type="spinal cord"
  /dev_stage="adult"
  /lab_host="DH10B"
  /note="Site_1: SalI; Site_2: BamHI; cDNA library was
  prepared and sequenced in Mouse Genome Encyclopedia
  Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in
  RIKEN, Division of Experimental Animal Research in Riken
  contributed to prepare mouse tissues. 1st strand cDNA was
  primed with a primer [5',
  GAGAGAGAGAGATCCAGAGACTCTTTTGTTTTATTTTNN 3'], cDNA was
  prepared by using trihalose thermo-activated reverse
  transcriptase and subsequently enriched for full-length by
  cap-trapper. cDNA went through one round of normalization
  to Rot = 10.0 and subtraction to Rot = 459.0. Second
  strand cDNA was prepared with the primer adaptor of
  sequence [5', GAGAGAGAGATCTCGAGTGAATTAATATACCCCCCCCCC
  3']. cDNA was cleaved with XhoI and BamHI. Vector: a
  modified pBluescript KS(+) after bulk excision from Lambda
  FLC I."
a 157 c 172 g 142 t

```

	151 a	157 c	172 g	142 t
BASE COUNT				
ORIGIN				

Query Match	9.5%;	Score	502;	DB	9;	Length	622;
Best Local Similarity	87.9%;	Pred.	No. 8e-93;				
Matches	547;	Conservative	0;	Mismatches	75;	Indels	0;
Gaps	0;						
<hr/>							
QY	36	gctccttttctgtcttctgccttgagttccagacacgcacgctccgaggagaccg	95				
Db	1	GCTGCTGGTTCTGGCTTTCCTACTCCTGAGTTCAGGACTGCACGCTCAGAGGAGACCG	60				
QY	96	ggacggcctatgggatgcctggggcccattgagtgaatgctcacgcacctcggggggtgg	155				
Db	61	TGAAGGCCCTCGGGATGCTTGCGGGCCCTTGGAGCGAGTGTTCCAGAACCCTGTGGTGGGG	120				
QY	156	ggcctcctactctcttgagcgctgctcagacagaagagctgtgaagaaagaataatccg	215				
Db	121	TGCTCTCTATTCCCTTGAGACGCTGCTCAGCAGCAAGAGCTGTGAAGGGAGAAATAITCG	180				
QY	216	atacagaacatgcagtaatgtgactgccaccagaacaggtgattccagctcagca	275				
Db	181	ATATAGAACAATGCAGTAATGTGGACTGCCACCAGAACAGGTGATTTCCGAGCTCACCA	240				
QY	276	atgctcagctcataatgatgtccaagcaccatggccagtttatgatgaatggcttcctgtgc	335				
Db	241	GTGTTCTTGGCTAATACGAGCTCAAGTACCATGSACAGCTTTATGAATGGCTTCCCTGATC	300				
QY	336	taatgacctgcacaaccaatgttcaactcaagtgcacagcccaaggaacaaacctggttgt	395				

Db	301	TAATGACCCGACAAATCCATGCTCACTCAAGTGCAGAAAAGCAACAGCGCTGTTG	360
QY	396	tgaactagcacctaaaggttttagatggttacggttgctatacagaatcttgatatgtg	455
Db	361	GGAACTAGCACCGAAGGTTCTTAGTGGTACGGGATGCTATACAGAATCACTGGATAGTG	420
QY	456	caicagtggtttatgccaaatgttgctgcataccacgctgggaagcacgcgtcaagga	515
Db	421	CATCAGCGGCCATGCGCAAAATGTTGGCTGTATCACACAGCTGGGAGGACACTGTCAACA	480
QY	516	agataactgtgggttcgaacggagatgggttcacctgcggctggttcgaaggcagta	575
Db	481	GGATACTGTGGAGTGTGCCACGGAGATGGCTCAACGTGCCGCTGGTCGAGGGCAGTA	540
QY	576	taaatcccgcttcgcgaacaaatcggatgatactgtggttgcaattccctatggaag	635
Db	541	TAAATCTAGCTCTCTGCAAGTAAATCGGATGACACGGTGGTCCCATCTCTACGGCAG	600
QY	636	tacacatatcgccttgcttta	657
Db	601	CACACACATCCGGCTGTGCTTA	622
RESULT	8		
LOCUS	AK020115	979 bp	linear
DEFINITION	Mus musculus 12 days embryo male wolffian duct includes surrounding region cDNA, RIKEN full-length enriched library, clone:6720426B09:Thrombospondin type 1 domain containing protein, full insert sequence.	HTC 19-JAN-2002	
ACCESSION	AK020115	1 GI:12860599	
VERSION	AK020115.1		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) 12 days embryo male wolffian duct includes surrounding region cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:6720426B09.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (sites)		
TITLE	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
JOURNAL	High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
AUTHORS	2 (sites)		
TITLE	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)		
PUBMED			

MEDLINE	20499374	
PubMed	11042159	
REFERENCE	3 (sites)	
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hagama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Todawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE	20530913	
PubMed	11076861	
REFERENCE	4 (sites)	
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
TITLE	Functional annotation of a full-length mouse cDNA collection	
JOURNAL	Nature 409, 685-690 (2001)	
REFERENCE	5 (bases 1 to 979)	

4 (sites)  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE  
Functional annotation of a full-length mouse cDNA collection  
JOURNAL  
Nature 409, 585-690 (2001)  
REFERENCE  
5 (bases 1 to 979)







sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

#### FEATURES

Location/Qualifiers  
 1..660  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="D330029102"  
 /clone\_lib="RIKEN full-length enriched, 13 days embryo heart"  
 /issue\_type="heart"  
 /dev\_stage="13 days embryo"  
 /lab\_host="DH10B"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
 GAGAGAGAGCGCGGCAACACGATGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
 GAGAGAGATCTCGAGTTAATTAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."  
 BASE COUNT 161 a 163 c 191 g 145 t  
 ORIGIN  
 Query Match 8 6%; Score 456.8; DB 9; Length 660;  
 Best Local Similarity 88.8%; Pred. No. 1.6e-83;  
 Matches 494; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
 QY 1 atggaatgctgcgtggcgaaactctggcacactgctctcttcttggtcttctgctc 60  
 Db 105 ATGGAATGTCGCGTGGCGACGCTCTGGCACACGCTGCTGTTCTGGCTTTCTACTC 164  
 QY 61 ctgaattccaggaccgcacgctccgaggaggacggcgccctatgggagctctggggc 120  
 Db 165 CTGAGTTTCAGGACATGACGCTCAGAGAGAACGCTGAAGGCGCTTGGGATGCTGGGCG 224  
 QY 121 ccatgagtgatgctcagcacctcggggggtggggctctctctctctctctgagcgctgc 180  
 Db 225 CTTGGAGCGAGTGTTCAGAACCTGTGGTGGGGTGGCTCTCTATTCCTGAGACGCTGC 284  
 QY 181 ctgagcagcaagagctgtgaaggagaataatccgcatcagcaacgacgagtaagtgcac 240  
 Db 285 CTGAGCAGCAGAGCTGTGAGGAGGAATAATTCGATATAGAACATGACGATGATGAGAC 344  
 QY 241 tccccaccagaagcaggtgattccgagctcagcagcaatgctcagctcataatgatctcaag 300  
 Db 345 TGCCCCACCAGAAGCAGGTGATTTCGAGGCTCAACAGTGTCTGCTCATACGACGCTCAAG 404  
 QY 301 caccatgcccagtttatgaatgcttctctgtgtctaaagccctgacacccatgttca 360  
 Db 405 TACCATGGACAGCTTTATGAATGGCTTCTGTATCTAATGACCCGACCAATCCATGCTCA 464  
 QY 361 ctcaagtgcacagcacaagcaacctggtgttgtaactagcacccttaaggtcttagat 420

Db 465 CTCAGTCCAGCAAAAGGACCAAGCCTGGTGTGGAACTAGCACCGAGGTCTTAGAT 524  
 QY 421 ggtacgctgtctatcacagaatctttgatgtgcatcagtggtttatgccaaattgtt 480  
 Db 525 GGTACGGGATGCTATACAGAATCACTGATATGTGCACTCAGGGCTATGCCAAATTTGT 584  
 QY 481 ggtcgtcatcaccagctgggaagcaccgtcaagaagataactgtgggtgtgcaacgga 540  
 Db 585 GGCTGTGATCACCAGCTGGGAGGACCTGTCAAGAAGATAACTGTGGAATGTGCAACGGA 644  
 QY 541 gatgggtccacctgcc 556  
 Db 645 GATGGCTCAACGCTGCC 660

#### RESULT 11

BB643318  
 LOCUS BB643318 RIKEN full-length enriched, 9.5 days embryo parthenogenote  
 DEFINITION Mus musculus cDNA clone B130031C01 5', mRNA sequence.  
 ACCESSION BB643318  
 VERSION BB643318.1 GI:16478052  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 700)  
 REFERENCE Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 UNPUBLISHED (2001)  
 CONTACT: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gs.riken.go.jp,  
 URL: <http://genome.gsc.riken.go.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

#### FEATURES

Location/Qualifiers

```

source
1..700
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="B130031C01"
/cloned_lib="RIKEN full-length enriched, 9.5 days embryo
parthenogenote"
/tissue_type="parthenogenote"
/dev_stage="9.5 days embryo"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5']
GAGAGAGAAGATGCCAAGAGCTCTTTTTTTTTTTTNN 3', cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of subtraction to
Rot = 229.0 Second strand cDNA was prepared with the
primer adapter of sequence [5']
GAGAGAGATTCGTGAGATTAAATAATTAATCCCCCCCCCCC 3']. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I.*
BASE COUNT      171 a   174 c   197 g   158 t
ORIGIN

Query Match          8.4%; Score 443; DB 9; Length 700;
Best Local Similarity 86.1%; Pred. No. 1.le-80;
Matches 502; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

Qy 1 atggaatgtcgccctcgggcaactcctggcacactgtctctcttcttggtttcctgc 60
Db 40 ATGGAATGCTGCCCTCGGGACGTCCTGGCACCGCTGCTGCCTTGCTTGGCTTCTACTC 99

Qy 61 ctgagttccaggaccgcacgtccgagaggaccggacgacctatggatgcctggggc 120
Db 100 CTGAGTTCAGAGACTGCACGCTCAGAGAGAACCCGTGAANGGCCCTCTGGGATGCTGGGGC 159

Qy 121 ccattgagtgaatgctcacgcacctcgcggggtgggcccctctactctctgagcgctgc 180
Db 160 CTTTGGAGCGAGTGTTCAGAACCTGTGGTGGGGGTGGCTCTTATTCCTCAGACGGCTGC 219

Qy 181 ctgaqcagaagagctgtgaaggagaataatccgatatacagaacaatgcagtaatgtgac 240
Db 220 CTGAGCAGCAAGAGCTGTGAAGGGAGAAAATTTCGATATAGAACAATGAGTAATGTGAC 279

Qy 241 tgccaccagaagcaggtgatttcagagctcagaatgctcagctcataatgatgtcaag 300
Db 280 TGCCACCAGAAAGCAGGTGATTTCGAGCTCAACAGTGTTCGTCTCATACGACGCTAAG 339

Qy 301 caccatgccagtttatgaatggcttcctgtgtctaataagccctgacaaaccattgtcca 360
Db 340 TACCATGACAGCTTTATGAATGGTTCTCTGTATCTAATGATGCCCGCAATCCATGCTCA 399

Qy 361 ctcaagtgccaaagcacaagcaaacccctggttgttaactagcacctaaaggtcttagat 420
Db 400 CTCAAAGTGCCAAGCAAAAGGAACCAAGCCGGCTGTGGAACTAGCACCCGAAAGGCTTAGAT 459

Qy 421 ggtaacggttgctatacagaatctttgatatgtcatcaggtggtttatgccaaattggt 480
Db 460 GGATCGCGATGCTATPACAGAAATPACTGGGATATGTGCATCAGCGGCTATGCCAAATGTT 519

Qy 481 ggctgcgatcacagctgggaagcacctgcaaggagaataaactgtggggtctgtcaacgga 540
Db 520 GCGTGTGATCACCACACTGGGAGACACTGTCAAAGAGGATAACTTGGGAGTGGGCAACGGA 579

Qy 541 gatgggtccacctgcgcgtggttcgaggggcagataaatccc 583
Db 580 AATGGCT-CACGGGCGGCTGGCCCGGACGAGATAAATTTTCAC 621

```

```

Db 209 TGGACAGAGTGTTCACAAAGCTGGGGGAAGGCACCCAGACTCGAAGCGGCATTTGCCGA 150
QY 2449 aagatgctgaaacccgctctcaacggttgcaactccacccctgtgccgcgcctgcct 2508
Db 149 AAGATGCTGAAACCGGCATCTCAACGGTGTGTCATTTCCACCTGTGCCGCCCTGCCT 90
QY 2509 ttcttctccctccatcagccctgtatgctggaacctgtgcaaggccggcgccatcc 2568
Db 89 TTCTCTTCTCCATCAGGCGCTGTATGCTGCAACCTGTGCAAGCGCGCGGCCCATCC 30
QY 2569 acgaagcacagcccgacatcgcgccgc 2597
Db 29 ACGAAGCACAGCCGCGACATCGCGGCGC 1

RESULT 13
AI750970
LOCUS
DEFINITION
AI750970 419 bp mRNA linear EST 22-JUN-1999
clone NHTBC_cn06h02 Random, mRNA sequence.
ACCESSION
AI750970
VERSION
AI750970.1 GI:5129234
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 419)
JIA,L.B., YOUNG,M.F., TOUCHMAN,J.W., BOUFFARD,G.G.,
BECKSTROM-STERNBURG,S.M., GREEN,E.D., POWELL,J.I., YANG,L.M., ROBAY
P.G., HOTCHKISS,R.N. and FRANCOMANO,C.A.
SGAP: The Skeletal Genome Anatomy Project
Unpublished (1997)
Contact: Libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-402-4877
Fax: 301-496-7157
Email: libin@helix.nih.gov
DNA Sequencing and analyses by National Institutes of Health
Intramural Sequencing Center (NISC).
Plate: 06 row: h column: 02
Seq primer: M13Rpl reverse primer (ABI).
FEATURES
source
location/Qualifiers
1..419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NHTBC_cn06h02"
/clone_lib="Normal Human Trabecular Bone Cells"
/sex="Female"
/tissue_type="Bone"
/cell_type="Trabecular Bone Cells"
/lab_host="SURE"
/note="Organ: Hip; Vector: pBluescript; Site_1: EcoRI"
BASE COUNT 111 a 112 c 114 g 81 t 1 others
ORIGIN

Query Match 7.8%; Score 414.8; DB 9; Length 419;
Best Local Similarity 99.3%; Pred. No. 5.8e-75;
Matches 416; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3503 gcaagatctcagccagccagctctcagctcggaggtggtcaccacccctggggcaga 3562
Db 1 GNACAGCTCAGCGGCCCGCAGCTCTCAGCCTCGAGGTGCTCACCCACCTGGGCCAGA 60
QY 3563 cgggtggccctgagcaggggacactgaggtgtctctcactgtgagggccatcgccacc 3622
Db 61 CGGTGGCCCTGGCAGCGGGACACTGAGTGTCTCTCTGCTGAGGCCATGGGCCACC 120
QY 3623 caagcctaccatcagctggcgccagggaatgagaagaagttcagttcagtagcaggattc 3682
|||||

```

```

Db 121 CAAGGCTTACCATTGAGCTGGGCCAGGAATGAGGAAGAGTTTCAGTTAGTCACAGGATTC 180
QY 3683 ttctacagccagatgattccttacagatctctggcaccagtggaagcagatgtgggtttct 3742
Db 181 TTCTACAGCCAGATGATTCCTTACAGATCTTGGCACCAGTGGAAAGCAGATGTGGTTTCT 240
QY 3743 acattgcaatggcacaatgcttggtgatacagactctgtctccattgctccacattag 3802
Db 241 ACATTGCAATGGCACCACCAATGCCCTGGGATACGACTCTGTCTCCATTGCCCTCATATTAG 300
QY 3803 caggaagccactagtgaacacgtcacgaatgacagtgatcaacacgagaaagcctgacg 3862
Db 301 CAGGAAAGCCACTAGTCAAAAGCTCAGCAATGACAGTATCAACACCGAGAGGCTGCAG 360
QY 3863 tcacagtcgatattaggaagcaccatcaacacagtcgaggagtggaatgtgacaatcaac 3921
Db 361 TCACAGTCGATATAGGAAGCACCATCAAAACAGTGCAGGAGTGAATGTGACATCAAC 419

```

## RESULT 14

```

BB478282 619 bp mRNA linear EST 25-OCT-2001
BB478282 RIKEN full-length enriched, 13 days embryo heart Mus
musculus cDNA clone D330029102 3', mRNA sequence.
BB478282
VERSION
BB478282.2 GI:16440420
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 619)
AUTHORS
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 23, 2000 this sequence version replaced gi:9395891.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

```

TITLE  
JOURNAL  
COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

1. 619

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="D330029102"

/clone\_lib="RIKEN full-length enriched, 13 days embryo heart"

/tissue\_type="heart"

/dev\_stage="13 days embryo"

/lab\_host="DH10B"

/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATATCCCGCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

160 a 142 c 169 g 148 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 7.8%; Score 413.8; DB 9; Length 619;

Mismatches 451; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 89 aggacgggaacgctatggatgctggggccatggagtgatgctacgacacgtcg 148

DB 1 AAGACCGTGAAGGCTTCTGGGATGCTTGGGGCCCTTGGCGAGTGTTCGAGACCTGTG 60

QY 149 ggggtggggcctactctctgagcgtgctgctgagcagcaagagctgtgaaggagaa 208

DB 61 GTGGGGTGTCTCTATCCCTGAGACGCTGCTTGAGCAGCAAGAGCTGTGAAGGGAGAA 120

QY 209 atatccgatacagaacatgcagtaattgtgagctgcccaccagcagcaggtgattccgag 268

DB 121 ATATTGGATATAGACATGCAGTAATGTGGACTGCCACCAGCAAGCAGGTGATTCGGAG 180

QY 269 ctgagaatgctcagctcataatgatgtcaagcaccatggccagttttatgaatggcttc 328

DB 181 CTCACAGTGTCTGCTCATACACGCTCAAGTACCATGACACGCTTATGCAATGGCTTC 240

QY 329 ctgtgttaataagccctgacaacccatgttcaactcaagtccaagcccaagggaacaccc 388

DB 241 CTGTATCTAATACCCGACAAATCCATGCTCACTCAAGTCCCAAGCAAAAGGAACACGCC 300

QY 389 tgggttgtaactagcactaaagttcttagatggtgaacgcttgctatacagaatctttgg 448

DB 301 TGGTTGTGGAACATAGCACCAGGCTCTTAGATGTGTAGCGATGTATACAGAAATCACTGG 360

QY 449 atattgcatcagtggtttatgtcaaatgttgctgctgcataccagctgggaagcaccg 508

DB 361 ATATGTGCATCAGCGGCTATGCCAAATTTGTTGGCTGTGATCACCACCTGGGGAGCACTG 420

QY 509 tcaaggagaataactgtgggtgtgcaacggagatgggttcacactgcgcgtggttcgag 568

DB 421 TCAAGAGGATAACTGTGGAGTGTGCAACGGAGATGGCTCAACGTGCGCGGTGTCGCGAG 480

QY 569 ggcagtatataatccagctctccgcaacccaaat 601

Db 481 GGCAGTATAAATCTCAGCTCTCTGCAAGTAAT 513

RESULT 15

BB615232

LOCUS BB615232

DEFINITION BB615232 RIKEN full-length enriched, adult male testis Mus musculus cDNA clone 4930443P21 5', mRNA sequence.

ACCESSION BB615232

VERSION BB615232

KEYWORDS BB615232.1 GI:16455520

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 555)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Atakawa,T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp

URL:<http://genome.gsc.riken.go.jp/>

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipillar sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Imanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

e mouse tissues.

Location/Qualifiers

1. 555

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="4930443P21"

/clone\_lib="RIKEN full-length enriched, adult male testis"

/sex="male"

/tissue\_type="testis"

/dev\_stage="adult"

/lab\_host="SOLR"

/note="Site\_1: XhoI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAAAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGAGGGCGCAATTAATTCGAGTAAATTAATCCGCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites. "

BASE COUNT 141 a 150 c 143 g 120 t 1 others  
ORIGIN

Query Match 7.7%; Score 409.4; DB 9; Length 555;  
Best Local Similarity 88.6%; Pred. No. 8.2e-74;  
Matches 443; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3513 agcgccagcagcagctcagcctcgaggtggtcaccacctgggacagcgtggccct 3572  
Db 1 AGCGGCACAGCAACTCTCAGCCTCGGAGGTAGTCACCCACTGGGCAGACTGTGGCTCT 60

QY 3573 ggcagcgggacactgagtggtctctgcaactgagggcatggccaccacagggctac 3632  
Db 61 AGCCAGTGGGACCTGAGTGTGTTCTGCACTGTGAGGGGTAGGCAACCCAAAGGCCAC 120

QY 3633 catcagctggccaggaatggagaaagtctcagttcagtgacagattctctacagcc 3692  
Db 121 CATCCACTGGACGACGAGCAAGCAGTCAATTCAGTGACAGGATTCCTACAGCC 180

QY 3693 agaatattcttacagatttggcaccagtggaagcagatgtgggtttctacacttgcaa 3752  
Db 181 AGATGATTCCTTACAGATCTTGGCACCTGGCAAGCTGATGGGTTCCTATACTGCAA 240

QY 3753 tgccaccatgccttgggatacgactctgtctccattgcccgtcacattagcagggaaagcc 3812  
Db 241 TGCCACAATGCTTGGGATATGACTCTCTCTCCATGTGTCACATTAGCAGGAAAGCC 300

QY 3813 actagtgaacgctcacgaatgacagtgatcaacacgagagagcctgcagtcacagtcga 3872  
Db 301 ACTAGTGAAGACATCCGAATCAGTCTGCTCAACACAGAGAGCCCTACAGTACAGTGA 360

QY 3873 tataggaagcaccatcaaaacagtcagggagtgaaatgtgacaatcaactgccaggttgc 3932  
Db 361 TATCGTGGCACCCTCAGAACCGTCCGAGGAGTGAATGTGCAATTAATGCAAGAGGC 420

QY 3933 aggagtcctgaagtcagtcacttgggtcaggaataaaagcaaaactgggtcccccga 3992  
Db 421 AGGTGTACCTGAAGCTGAAGTCACTGGTTCAGAAATTAAGCAAAATTTGGGCTCTNCCA 480

QY 3993 ccattgcacgaaggtcct 4012  
Db 481 CCACCTGATGAGGCTCCT 500

Search completed: July 24, 2002, 00:27:47  
Job time: 6781 sec

